

Making headway for science

The president of the Royal Society has spoken out publicly in protest at what is happening to British science. The hope must be that his modest demands are granted — and that it is not too late.

SIR George Porter, president of the Royal Society, is not a particularly subversive man. He has not, so far as is known, marched through the streets of London in protest at a government decision or joined a mass lobby at the House of Commons in the hope of getting some policy changed. For all that the world knows (and if it were relevant), he may even be among the British government's most loyal supporters when its periodic re-elections come around. It is therefore all the more pleasing that he should have used the occasion of the BBC's annual Dimpleby lecture last weekend to tell the British government in plain language what must be done to put some stuffing back into British science. The message should be all the more memorable (in the minds of those prepared to listen) for being barbed as well as amusing.

The essence of the case, spelled out more soberly in last November's anniversary address to the Royal Society and elsewhere, is that there is nothing wrong with the wishes of governments to turn science into prosperity, but that the British government does the job badly because it does not collectively understand that scientific discovery is the product of able people's personal curiosity or appreciate the distinction between research and technological innovation. A telling object lesson will be found below. (Porter has obviously found civil servants even more insensitive than ministers; he took a previous permanent secretary at the Treasury explicitly to task for telling a House of Lords committee that "the national source of science and technology is less important than the ability to assimilate . . . ideas whatever their origin".)

Few will dissent from the ingredients of the argument. The precise degree to which the British research enterprise has been damaged in the past few years may be disputed, but there is no denying that "once a base of knowledge and skill has been removed . . . it is virtually impossible to catch up with the uninterrupted competition". While acknowledging the need that governments should "take care of the pennies", Porter is surely on firm ground in complaining at the government's failure to recognize the importance of the scientific enterprise. ". . . we hear frequent speeches emphasizing the importance of exploiting science", but support for science itself elicits "the enthusiasm of an atheist supporting the church".

The distinction between research and technological innovation is well put (with the help of the definition of the Royal Society's own purpose as "the improvement of natural knowledge"); the pursuit of knowledge and its useful application are indeed distinct, although not antithetical, activities. Research does depend on young people (which does not mean older colleagues have no part to play), the support the able among can command and the freedom they are given to satisfy their own curiosity. Porter is right to emphasize the importance, in present-day Britain, of the difficulties faced by young people seeking support for their research. (See, for example, the story on page 578.)

Exploitation needs different skills (among which a knowledge of the market and of a company's strategy may be crucial) and entails the spending of money by industry (on a scale probably larger than that to which British industry is accustomed). Porter

rightly ridicules the British government's standard misconception that, if British industry has historically failed to capitalize on British science, the fault must lie with the academic partners. He is also right (and courageous as well) to point to the irony that, in a Britain in which people clamour for the application of science, Members of Parliament will later this year have a free vote on the government's modest proposals for embryo research, hedged about though they are with regulations and restrictions of the kind suggested by the Warnock Committee set up to look into this issue some years ago. How, Porter's question goes, can you hope profitably to exploit new knowledge if those who would gather it are impeded from doing so by the law as well as the lack of funds?

The argument is compelling and, curiously, may make some headway for having been made on television. But it is also plain that the British have worked themselves into a mess about science and public support for it. It is difficult to tell whether the government's inflexibility stems from an erroneous theory of the relationship between science and technology or from simple animus towards academics and academic institutions. The best hope is that changing circumstances will allow people to change their views without appearing to suffer defeat. There are signs of change — not just the occasional words of approval that ministers drop at scientific occasions. The most promising is the improvement of the British economy — the improvement not merely of the government's finances but also of companies' capacity to invest in the future (which may yet allow them to hire more technical people or even to pay those they have more sensibly). Porter last weekend asked for a gesture of good faith — £25 million for equipment grants to able people whose applications have been refused, and £2 million a year to make research grants too small for the research councils' administration. It is a modest demand, but one that the government can hardly refuse. □

Exploitation gone awry

British efforts to turn information technology into prosperity have not come to much.

ALL governments seek ways of turning science into prosperity, but only two have had much success; that of France, which over three decades has been sustained in that cause by the tendency to wrap technology in the tricolour, and that of Japan, whose Ministry of International Trade and Industry (better known as MITI) has the good sense to be the perpetual chairman of a committee of market-driven manufacturers, not a puffed-up strategist. The British government has a conspicuously poor record in the field: the then Mr Harold Wilson's Labour government earned so little credit for its promise in 1964 of a "white-hot technological revolution" that its successors have mostly (and wisely) said it is their job merely to create the circumstances in which innovation can flourish. But the old interventionist Adam, it seems, is not easily restrained.

In 1982, throwing its own principles to the winds, the British government embarked on a £350-million five-year programme



of research in information technology. Last week, the National Audit Office (NAO) published an interim report (see p.579) on what is called the Alvey Programme (after five years, still halfway through its term) which, if it were not such sad proof of governments' recurring vulnerability to conceit, would make an amusing read.

Learning skills

The circumstances are these. In March 1982, Mr John Alvey (then and still a director of the telecommunications business called British Telecom) became chairman of a committee to advise the British government on how to meet the challenge of the 'fifth generation' computer advertised the previous year by MITI (and since, apparently, forgotten). Alvey had been asked to take on this task by Mr Kenneth Baker, then Minister of Information Technology, now the much more important Secretary of State for Education and Science, who had spent much of his time preaching the need that a then-backward industrial country should learn new skills. Alvey can in these circumstances have had little choice but to recommend government sponsorship of some kind.

Even in retrospect, and despite the sarcasm with which the NAO report is laden, nobody can complain that the programme put forward was unintelligent, or pointless, or even extravagant. Alvey and his committee, Baker and Mr Brian Oakley, the administrative head of the Science and Engineering Research Council (SERC) who became director of the Alvey Programme, have laboured selflessly and diligently to make their enterprise succeed. Much good has incidentally come out of it. So why has the project also been a disaster?

At the outset, the idea was straightforward. Industry (software houses, chip manufacturers) and academics collaborate too little? Then use government funds to tempt them together. Britain is full of bright ideas, but nobody bothers to apply them? Then make sure that Alvey projects are geared to the marketplace. The output of graduates in computer science and electronic engineering from higher education is too small? Then create extra posts and places. Too many government departments have a finger in the pie already? Then set up an overarching system, called a directorate, to manage the work. The Alvey proposal was that the government should put up two-thirds of the total cost of £350 million, paying the whole cost of the development of higher education and of research by academic partners in Alvey collaborations and up to 90 per cent of the cost of research by industrial partners. Being experts, the Alvey committee also specified the fields in which spending should be concentrated — software engineering, Very-Large-Scale Integration (chip-making), Man-Machine Interface and Intelligent Knowledge-Based Systems (AI to some).

The government's first mistake was to applaud the Alvey message but to decline to pay the bill. Are we not already spending £5,000 million a year on research and development? It might truthfully have asked itself (and its then much less influential member, Mr Baker). Instead the funds were quarried from the existing budgets of other agencies — SERC, the Ministry of Defence and Baker's then ministry, the Department of Trade and Industry (DTI). One direct result of that parsimony is that payments to different members of Alvey collaborations have come from different sources and on different contractual terms, to the general confusion. SERC, persuaded at the outset to commit £10 million a year to Alvey projects, had had to spend more than £75 million by last summer and has at least a further £5 million to find. At the same time, the cost to the industrial partners will be less than originally foreseen, chiefly because the Alvey directorate has had to spend £35 million more on infrastructure and administration than the nil provision in its budget. It is hard to believe that SERC's hard-pressed managers and their academic constituents will now be overjoyed that the Alvey director should have told NAO that this shortfall may not affect the work done because academic

costs are lower.

NAO also describes how the programme was confused, and individual projects often seriously delayed, by the difficulty of persuading partners in collaborations to make legal agreements codifying their relationships. One obstacle seems to have been the unfamiliarity of the exercise, that of arranging collaboration in what is vaguely defined as 'pre-competitive' research, but there were also serious problems about the ownership of intellectual property. None of this is surprising. Everybody was working in the dark. NAO attributes to the Alvey directorate the opinion that its experience will help pave the way for smoother future collaborations between industry and academics, but that seems a troublesome and expensive way of learning some simple lessons.

Even the ambition that the Alvey Programme should draw smaller companies into collaborations seems to have been frustrated by bumbledom at the centre. The original proposal that the government should pay up to 90 per cent of industrial project costs was intended to help small companies take part, but in the event the limit was fixed at 50 per cent, with the result that in mid-1986 the five largest participating companies accounted for nearly half the work. The Alvey directorate may say (as quoted by NAO) that this merely reflects the fact that larger companies do the most research, but one objective of the exercise was to redress that balance.

NAO's criticisms are explicitly concentrated on the machinery of the programme, not its content. For what it is worth, there seem to be many centres involved in Alvey projects at which constructive work appears to have been well done, the uncertainties and confusions notwithstanding. In one respect, NAO goes unreasonably far, by drawing attention to the small number of Alvey projects that have so far led to marketable products. Even the Alvey committee's original three-year yardstick for successful implementation was optimistic; now to judge projects by the test of whether they succeed within a year is unwise. It is also important that the difficulties NAO catalogues are probably unavoidable when governments decide that a committee's scheme must be implemented quickly, before those asked to do the work have time to think through the difficulties ahead, and on a shoestring.

Shadow

That is the sense in which NAO's report casts a shadow over whatever lies ahead. Britain will hear a lot more of collaboration between academics and industry. Earlier this year, the department of Trade and Industry named itself the 'department for enterprise' and, while eschewing 'near-market' research, talked of its ambitions to foster other schemes along Alvey lines. Will it learn from NAO's report, and if so, what? Will it, as it might well, go so far as to understand that many of the difficulties encountered by the Alvey Programme reflect the underlying error of seeking to harness the interests of academics to goals of a near-commercial character defined not by themselves, but by a committee? What is to be said about the even larger collaborative research programmes (the largest also in the same field) run by the European Commission? And whatever is to happen to the continuation of the Alvey Programme, on which a committee under Sir Austin Bide pleaded more than a year ago that more money should be spent? NAO puts that story laconically. "The government welcomed the report . . . said its recommendations were being considered . . . DTI has provided £29 million over the next three years and SERC £55 million over a five-year planning horizon". The lesson the British (and other) governments should take to heart is simply that their original instincts are correct. Governments may properly seek to create the climate in which innovation prospers, but are ill-equipped to manage innovation and liable to seem incompetent when they try. The pity (for British taxpayers) is that the history of the past forty years is littered with object-lessons with the same point. Will governments ever learn? □

Nerve gas cloud hangs over West German firms

- Nerve gas link between Iraq and Germany
- Companies face severe penalties

Munich

DID West German companies ship nerve gas production facilities to Iraq? The question is being asked again in the wake of Iraq's apparent use of chemical weapons in an attack on the Kurdish village of Halabja in northeastern Iraq on 17 March.

Uproar over the issue caught a slumbering Bonn by surprise. The West German capital was nearly empty last week for the Easter parliamentary recess. But opposition leaders demanded an investigation to determine if West German companies had delivered factories capable of producing nerve gas. Foreign Minister Hans-

Dietrich Genscher (Free Democrat) called for an international ban on the production of chemical weapons.

Iraq is reported to have used the nerve gases tabun and sarin, derivatives of materials used in pesticides and fertilizers. The United Nations has claimed repeated use of chemical weapons by Iraq against Iranian troops since 1984.

West German law forbids the shipment of facilities that might be used to produce chemical weapons. This would include the pesticide factories described above. Criminal penalties for such export include gaol terms of up to three years and unlimited fines. The penalties for exporting chemical or other weapons themselves to Iran or Iraq are far harsher. The use of chemical weapons violates the 1925 Geneva Protocol banning their use. Both Iran and Iraq are signatories of the agreement.

Twelve companies suspected of having violated the law are already being investigated. Raids on the companies' offices in November 1987 yielded thousands of documents.

One company being investigated, WET (Water Engineering Technology) of Hamburg, admits that it shipped up to 70 per cent of the parts necessary for a pesticide and fertilizer factory to Iraq between 1985 and 1987, when the shipments first drew public notice. WET's lawyer, Thomas Marx, also admits that the pesticides could be used as precursors to chemical weapons "with five or six additional steps". But Marx says that the parts were never assembled.

A spokesman for the Foreign Ministry said that the government has been diligent in updating the laws covering export to Iraq and Iran. A clause was added to the foreign trade laws as early as 1984 requiring licences for the export of facilities that could create the precursors for chemical weapons. No licences have ever been issued. The laws were strengthened in January 1987 to include more possible precursors in accordance with a European Community measure.

Most sources agree that the larger problem — that Iraq has the ability to produce poison gas and is not afraid to use it — will not be solved by West German action alone. Iran has threatened to retaliate in kind if the attacks do not stop. It remains to be seen whether Genscher's initiative comes soon enough, and is strong enough, to prevent escalation. **Steven Dickman**

US and Soviet science deal to be renewed

Washington

THE United States and the Soviet Union will begin work in earnest next month to renew the basic science agreement between the two countries that lapsed six years ago. Other scientific agreements in the fields of transportation and the environment also are receiving renewed attention, and even the controversial issue of AIDS (acquired immune deficiency syndrome) may soon receive direct government-to-government attention.

Scientific agreements with the Soviet Union have been popular items in the United States lately. Earlier this year, the National Academy of Sciences renewed and expanded its agreement with the Soviet Academy of Sciences (see *Nature* 331, 197; 1988), and agreements on space and health have been invigorated (see *Nature* 323, 482; 1986 and 326, 234; 1987). The basic science agreement, designed to be an umbrella for all other scientific exchanges, was among those allowed to lapse in the early 1980s as a protest against the Soviet invasion of Afghanistan.

The US State Department and the Soviet Foreign Ministry have already held preliminary discussions about renewing the basic science agreement. Although scientists in both countries have been enthusiastic about renewed cooperation, there are still areas where the two governments do not see eye-to-eye.

Intellectual property rights and technology transfer issues are potential stumbling blocks. The United States is also keen to keep human rights questions on the table for discussion, insisting that any Soviet scientist should be eligible to participate in exchange activities.

Although both countries have been enthusiastic about cooperation in health, AIDS has remained a sore point. The United States took great exception to suggestions made in the Soviet press that the AIDS virus was part of some US biological weapons experiment gone awry, and refused to discuss anything about the disease until the Soviet Union repudiated such charges.

That appeared to have happened last year, and assistant secretary of health Robert Windom scheduled a visit to Moscow to discuss AIDS. But earlier this year press reports about the biological weapons scenario surfaced again, and the Windom trip was called off. After repeated assurances the United States is satisfied that the Soviet Union has abandoned its claims, and the Windom trip is on again, although no date has been set.

Joseph Palca

Home test kits for AIDS blocked

Washington

PLANS to market home kits to test for the human immunodeficiency virus (HIV) which causes AIDS have been quashed by new guidelines from the US Food and Drug Administration (FDA).

According to the FDA, more than twelve companies had expressed interest in selling test kits over the counter at drugstores and supermarkets in much the same way that pregnancy test kits are marketed. The kits would permit users to take a blood sample for mailing to a testing laboratory.

In a letter sent out to potential manufacturers of the kits last week, FDA stopped short of a ban on home testing. But such stringent criteria for pre-marketing research of home test kits are laid out in the letter that it is unlikely that anyone will persist in seeking approval.

The key provision in the FDA rules is that "a professional health care provider" must be responsible "for reporting and interpretation of the result to the requester of the test, as well as for counselling". FDA officials say they are frightened of the psychological damage that might be caused to someone who learns that he or she is HIV-positive through an impersonal letter from a testing laboratory, particularly if there is any possibility of a false-positive result.

Potential manufacturers expressed annoyance at the letter, arguing that it ran counter to the idea of freely available mass testing. An FDA spokesman said it will continue to gather information on the acceptability of home testing and may reconsider later. **Alun Anderson**

Mainstream scientists confront unorthodox view of AIDS

Washington

PETER Duesberg's quest for scientific credibility for his unorthodox theories about the causes of AIDS lost ground on Saturday during a scientific forum on the aetiology of AIDS sponsored by the American Foundation for AIDS Research. Duesberg has challenged the idea that human immunodeficiency virus (HIV) causes the disease, and has outraged health-care workers by suggesting that AIDS patients abandon treatment with drugs meant to control the virus, and by claiming that infected heterosexuals need not use condoms or change their sexual practices.

Duesberg, a well-known virologist and professor of molecular biology at the University of California at Berkeley, has achieved a certain notoriety in the past year from his ideas, which originally appeared in the journal *Cancer Research* (47, 1199; 1987). They have been largely ignored by the scientific community, and have gained credibility mainly among patient populations with the greatest interest in learning that HIV infection does not lead inevitably to a fatal disease. But many researchers now feel that Duesberg's arguments must be confronted, as the consequences of acting on his advice may be fatal.

Because HIV does not seem to infect a high proportion of cells, Duesberg has claimed that it cannot cause the widespread changes in the immune systems of AIDS patients. He points out that most viral diseases appear before an antibody response is mounted, whereas AIDS symptoms always appear after antibodies to HIV have developed. He also raises issues about the latency of disease following infection, the apparent preference of the virus for homosexuals, and the difficulty in isolating HIV from all AIDS patients. Risky behaviour, not HIV, is the problem, according to Duesberg.

All of these points were disputed by the researchers present. William Haseltine of the Dana Farber Cancer Institute noted the distinction between antibodies that neutralize the virus and those that merely recognize it. Because the immune system can 'see' the virus does not mean that it can fight it off. Harold Ginsberg of Columbia University, who acted as moderator for the forum, explained that even when neutralizing antibodies are present, disease can still occur, as is the case with polio, measles and hepatitis-B virus.

Anthony Fauci, director of the National Institute for Allergy and Infectious diseases, pointed out that HIV can be shown to be cytopathic *in vitro*, and that there are

many ways in which it could orchestrate the slow decline in the immune system. He disparaged the idea that behaviour alone could cause AIDS, asking what risky behaviour a newborn infant could be said to have.

For his part, Duesberg seemed unwilling or unable to answer these potentially fatal objections to his theories and did not explain the relevance of risky life style to AIDS in Africa or to animal models of AIDS. Duesberg's behaviour theory also fails to explain why AIDS appeared only in the past decade. The scientists assembled at the forum appeared nonplussed that someone who professes unfamiliarity with modern immunology would choose to make such sweeping pronouncements about a disease that involves such complex and subtle changes in the immune system.

Duesberg's theories will no doubt continue to receive attention from groups already mistrustful of the scientific establishment's response to AIDS. But scientific acceptance for his ideas about AIDS, never very high, seems to be sinking.

Rebecca Ward

High magnetic fields

WEST Germany and France agreed last month to renovate and expand the high-field magnetic laboratory (HML) in Grenoble, France.

The laboratory is available to all researchers for the investigation of materials under extremely high magnetic-field conditions. Its current capacity of 30 Tesla (a European record) will be increased to 35 Tesla over the next few years through grants of DM12.5 million (about \$7.5 million) from both the West German Ministry for Research and Technology (BMFT) and the French Centre Nationale de la Recherche Scientifique (CNRS). HML has been operated jointly by the two countries since 1963.

S.D.

China syndrome

CHINA is to have a new 'special scientific zone' in the Zhongguancun area of the Haidian district of Beijing. Like the special enterprise zones (SEZs) already established in coastal regions, the new zone will offer special incentives to foreign partners. Although several of the existing SEZs include science parks and high-technology industries, the location of the new zone has been officially hailed not only as a new breakthrough in China's 'open' policy but also as the answer to China's hopes of its own 'silicon valley' (or 'silicon paddy') in the near future.

V.R.

Soviet scientific set-up attacked

London

THE Soviet Academy of Sciences, regarded in the West as a bastion of democracy, is failing to respond to the challenge of *perestroika*, according to a sharp attack by Professor Maxim Frank-Kamenetskii in *Literaturnaya Gazeta*. The article appeared on the eve of last month's Annual General Meeting (AGM) of the academy.

Frank Kamenetskii, who is a section head at the academy Institute of Molecular Genetics in Moscow, contrasts the open discussions of *perestroika* among professional groups such as writers, filmmakers and theatre workers with the complacency at the academy.

He complains that at last year's AGM, the proceedings followed "the same old scenario", with the vice-presidents simply reporting on "further successes".

Unlike factory workers who can now vote for their managers, Frank-Kamenetskii says, scientists employed by the academy have no part in decision making. Two years after the "reorganization" of the academy, interference by directors in the work of laboratories is still the norm. And the accreditation commissions introduced in 1986 to assess the work of scientists are, in practice, nominated by the director and serve simply to endorse his own decisions.

Frank-Kamenetskii attacks, either openly or by implication, several deep-rooted features of Soviet scientific life: the false distinction of "our" science and foreign science, the poor quality of much Soviet published work and its poor showing in international citation indices.

Interestingly, though, the article rejects one of the "solutions" widely canvassed in discussions of recent educational reforms — the need to build up a research base in the universities and higher education institutes. Certainly, he says, the deficiencies of university research are a fault, but the victim is higher education, not research.

Frank-Kamenetskii argues that research establishments need not be vast concerns. The key, he says, is laboratory independence. Accepting that institutes may have to stand on their own feet, he asks that they be able to seek support from home and abroad and that they should be freed from what he calls the "scientific mafias", preferably by requiring that the administrators of research funds should serve only for limited periods.

Startlingly, Frank-Kamenetskii also urges that the appraisal of projects should include overseas referees. Anticipating the objection that this would lead to the premature leakage of discoveries, he says that it will not otherwise be possible to raise standards fast enough.

Vera Rich

Foreigners fear superconductor club is beyond their means

Tokyo & Washington

At the opening ceremony of Japan's International Superconductivity Technology Center at a Tokyo hotel in January, only a handful of Westerners were present. And it seems that the centre is going to stay that way. For, despite an open invitation to join, very few foreign companies think that the million-dollar membership fee makes it worthwhile.

The centre was established by private industry at the urging of the Ministry of International Trade and Industry (MITI) to carry out research and disseminate information on the new high-temperature superconductors. From the outset, international cooperation was invited, and there was much talk in Japan of avoiding the brutal competition that soured US-Japan relations over semi-conductors.

Full membership of the consortium is expensive. To use the centre's research laboratory requires a down-payment of ¥100 million (\$800,000) and an annual fee of ¥12 million (\$100,000). And that is just a start: then come research costs and, for foreign firms, the cost of sending staff and providing them with local back-up.

For a smaller annual fee of ¥2 million (\$16,000), companies get access to an information service, a newsletter and international symposia.

The laboratory should be complete in the autumn and will be headed by Professor Shoji Tanaka, recently retired from Tokyo University. Although more than 100 Japanese companies, including banks and security companies, are subscribing to the information service and 45 have joined the laboratory, foreign companies are not so keen. IBM Japan, Rockwell International and Du Pont, all of which have Japanese subsidiaries, will join the information service, according to a MITI official. But no foreign company has agreed to join the research laboratory.

Professor Roland Schmitt, president of Rensselaer Polytechnic Institute, New York, and chairman of the National Science Board, stresses that Japanese officials should not be over-hasty in blaming US companies for lack of enthusiasm for international cooperation. He points out that the original invitation to join the centre came with just a four-page explanation. "Any company research director would have been out of his mind to send an \$800,000 dollar cheque based on the information given", he says.

There are many opportunities to spend money on superconductor research in the United States, he points out, and there is a proven value in projects at US laboratories. And multi-company cooperative research institutes — a rarity

even in Japan — are not, he believes, such an important part of the Japanese success story that US firms must rush to join in.

Westinghouse Research and Development's Professor John Hulm says that his company has made no decision yet. He worries about the economics of joining the research centre and about whether researchers who do not speak Japanese could really hope to participate fully. Unfortunately, experts in high-temperature superconductors who also speak Japanese are rare! Du Pont's Dr Alan Lauder says his company will join the information centre and "hasn't precluded" joining the research centre.

Organizers of the centre are putting on a brave face. Risaburo Nezu of MITI's Agency of Industrial Science and Technology says that seven or eight foreign companies, including American Telephone and Telegraph (AT&T), are "thinking" about joining the laboratory. Nezu is also very

encouraged by a meeting he had two weeks ago with officials of Britain's Department of Trade and Industry. He felt strongly that Britain wants to cooperate and he is very "hopeful" that British companies will participate in the centre.

Tanaka appreciates US difficulties. He sees many problems to overcome, particularly in relation to patents and licences. According to Nezu, patents taken out in the centre's research laboratory will be jointly owned by the centre and the member companies which belong to the laboratory, but this must be examined "from the legal point of view".

At an international conference on superconductivity held in the United States at the end of February, Tanaka says, he approached one researcher after another without success. British researchers, on the other hand, have shown a very positive response, according to Nezu. And both he and Tanaka see a good possibility that two or three researchers will come to the laboratory from Britain. But a tiny fraction of foreign researchers will hardly justify calling the centre "international".

David Swinbanks & Alun Anderson

High-critical-temperature superconductor made from glass

Tokyo

High-critical-temperature superconducting glass? Not quite. But Japanese researchers at the Technological University of Nagaoka have succeeded in making glass



Making superconducting glass.

from the ingredients of various bismuth-based high-temperature superconductors. And subsequent annealing at high temperature (820°C) converts the glass into a superconductor. The development could be used in the manufacture of very fine superconducting wires.

The glasses are prepared by melting various combinations of bismuth, calcium, strontium, aluminium, lead and copper oxides at 1,150°C. The melts are then quenched by pouring onto an iron plate and quickly pressing them into a slab.

Samples of $\text{Bi}_{1-x}\text{Ca}_x\text{SrCu}_2\text{O}_y$ and $\text{Bi}_{1-x}\text{Ca}_x\text{SrCu}_2\text{O}_y$ prepared by the technique exhibit perfect glass behaviour — they are completely amorphous and when slowly heated between 400 and 500°C show an endothermic glass transition followed by exothermic crystallization. Melt-quenched samples containing lead or aluminium are also glass-like, but X-ray diffraction reveals small crystalline peaks.

Similar melt-quenching methods have already been applied to the preparation of high- T_c yttrium-based copper oxides but the melt-quenched yttrium ceramics are microcrystalline (for example, see *Nature* **330**, 511; 1987). This is the first report of glass, and the Nagaoka researchers call their new materials "high- T_c superconducting glass-ceramics".

The Nagaoka team has found that rapid cooling in liquid nitrogen after annealing produces superconductors with the highest transition temperatures. The highest T_c obtained so far is 78 K for $\text{BiAl}_{0.5}\text{CaSrCu}_2\text{O}_y$.

The great advantage of the glass ceramics is that they can easily be formed into various shapes. And Professor Kazumasu Matusita, who heads the research, says that it should be possible to draw the glass out into very fine wires using the same technology that has been developed to make optical fibres. Details of the glass preparation method will be published in the *Japanese Journal of Applied Physics*.

David Swinbanks

Sanction plans anger Japan

Washington & Tokyo

THE promise of tough action by the US Congress against the Japanese Toshiba group and the Norwegian company Kongsberg Vaapenfabrikk has evoked a strong protest from Japan, which fears that US attempts to police world technology will encroach on its independence.

Toshiba Machine Company and Kongsberg Vaapenfabrikk drew US anger by transferring advanced propellor milling machinery, developed in Japan, to the Soviet Union in defiance of regulations set by COCOM, the 16-nation Coordinating Committee for East-West Trade Policy set up to stem the flow of strategically sensitive goods to communist countries.

Both companies have been punished by their own governments. Toshiba Machine being hit by a one-year ban on sales to communist countries that will lose it around \$40 million. Now there is a possibility that they will be punished further by the US government, which Japanese and European diplomats complain will violate the sovereignty of COCOM countries and set an unwelcome precedent.

The punishment is proposed in the much-debated omnibus trade bill, which is finally in a form that both House of Representatives and Senate committees agree upon. If it is voted into law it will ban for three years all sales in the United States by both Toshiba Machine Company and Kongsberg Vaapenfabrikk. The Japanese parent corporation, the giant Toshiba Electronics Corporation, will also be punished with a three-year ban on sales to US government agencies. That ban could cost Toshiba Corporation alone some \$100 million a year.

Japan's International Trade and Industry Minister, Hajime Tamura, described the congressional measure as "foolish" and harshly rebuked US attempts to undermine the basic COCOM philosophy that each member nation should impose export controls voluntarily.

The bill provides a 2-5-year ban on US imports from companies that offend against COCOM regulations, whatever country they are located in. And it provides a retroactive punishment for those that have offended in the past, if they are linked to the Toshiba case. French and West German firms are under suspicion.

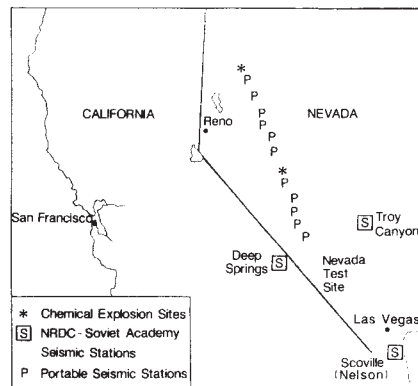
The trade bill contains several measures that offend US President Ronald Reagan, among them the punishment for Toshiba Corporation, and a presidential veto has been threatened if the bill passes through Congress without revision. Action on the bill in the House of Representatives is expected to begin this month.

Alun Anderson & David Swinbanks

Soviet and US scientists go for a blast in the Nevada desert

Washington

THERE are unusual goings-on in the Nevada desert. Earlier this year, the US Department of Energy hosted a team of Soviet scientists at the highly secure Nevada Test Site where the United States tests its nuclear weapons (see *Nature* 331, 380; 1988). Now another group of scientists from the Soviet Institute of Physics of the Earth will join US colleagues to detonate two chemical explosions at the site as part of an initiative sponsored by the



Natural Resources Defense Council (NRDC) and the Soviet Academy of Sciences to demonstrate technology for verifying a nuclear test-ban treaty.

The Soviet team, led by Senior Staff Scientist Nicolai Yukhnin of the Institute of Physics, arrived in the United States last week. They brought with them 1 ton of recording equipment to supplement the seismic monitoring devices already at the three NRDC-Soviet Academy stations

located in Nelson, Nevada, Troy Canyon, Nevada, and Deep Springs, California, on the Nevada border (see map).

According to current schedules, 10 tons of chemical explosives will be set off on 29 April at Black Rock Desert, and 15 tons will be detonated the following day at Broken Hills. Portable seismic stations will be set up to record the waves after the explosion. In both cases the land is owned by the US government.

The NRDC and the Soviet Academy first agreed to establish seismic monitoring stations around the Soviet and US nuclear tests sites in May 1986 (see *Nature* 321, 638; 1986). The first stations were placed in the Soviet Union in September 1986 about 200 km from the Kazakhstan test site and the US stations were established the following winter. Last September, the Soviet Academy detonated three chemical explosions at the Kazakhstan test site. Based on seismic data collected from those explosions, NRDC reckons some 25 stations in the Soviet Union and 15 in neighbouring countries would be needed to verify compliance with a treaty that limited underground nuclear test yields to one kiloton.

This is the fourth time Soviet scientists have visited the United States as part of the NRDC-Soviet Academy activities, but it is the first time that the State Department has not placed restrictions on their travel. A State Department official refused to characterize the unrestricted visas as representative of any change in US policy.

Joseph Palca

Evolutionary ships passing in the night?

Oxford

OXFORD University's Sheldonian Theatre was the venue for a piece of real theatre on Saturday when Richard Dawkins and Stephen Jay Gould held a "Darwinian Dialogue".

Gould revels in pluralist approaches to explaining organic diversity. He argues for a hierarchical view of life, with selection acting in the emergent properties at each level of the hierarchy and sorting the units at other levels. It may be unsatisfactory to interpret speciation rates just as properties of genes. Gould also stresses the importance of developmental constraints and historical contingencies, like mass extinctions or selection for one trait (such as brain size) resulting in many non-adaptive epiphenomena (like Bach's B-minor mass).

Dawkins, on the other hand, is a self-professed monist fascinated by organized adaptive complexity, who focuses on the gene as the unit of selection. He coun-

ters Gould's hierarchical view by distinguishing between replicators and vehicles. Genes, he argues, are the only true replicators. Dawkins concedes that the evolution of evolvability is another problem that interests him. Some embryologies may have become more widespread because they permit a wider range of phenotypic possibilities and so are good at evolving. He cites segmentation as an example.

Both started by proclaiming a dialogue rather than a debate since, as Gould pointed out, truth is only one of many weapons in a debate. After separate 30-minute presentations, the dialogue (or was it a debate?) started. Nothing was resolved. Whether gene versus organismal selection is a matter of semantics (Dawkins) or not (Gould) remains an open issue. But it was not clear, to us at least, whether they were attempting to answer the same question. Nonetheless it was all jolly good fun.

Andrew Pomiankowski & Paul H. Harvey

UK physics higher education "at severe risk" says institute

London

IN common with most other areas of science in Britain's higher education system, physics is in a serious state of decline, according to a survey by the Institute of Physics whose findings were released last week. The report of the study appears as the University Grants Committee (UGC) is conducting a detailed review of university physics, which is likely eventually to recommend a redistribution of resources, as was the case with the recently completed Earth sciences review.

According to the institute, financial stringencies have resulted in:

- a 20 per cent decrease in the number of academic staff in university and polytechnic physics departments over the past six years, with pressure for further reductions;
- a damagingly imbalanced age-structure of academic staff in physics departments, with fewer than one in seven under 35;
- the closure of physics departments in four universities since 1984 and the merging of physics departments in most polytechnics into departments of combined sciences or engineering;
- a growing problem of equipment obsolescence and shortage of technical staff in physics teaching laboratories.

The quality of physics higher education is consequently "at severe risk".

While the report concedes that a degree of concentration of equipment and resources is needed, it warns that Britain needs no fewer than 35 well-structured physics departments, each with at least 18 UGC-supported staff and 180 full-time equivalent students. If closure of departments extended to those that at present have between 18 and 25 UGC-funded staff, half of Britain's universities would be without a physics department, which, says the institute, would be "a calamity".

The study draws attention to the rising student-to-staff ratio, which is likely to continue. Between 1980 and 1986, the number of academic staff in physics departments fell by 17 per cent, while during the same period 21 per cent more students

graduated. According to the report, pressure from university administrators to meet planning targets is continuing, which will result in an estimated further 20 per cent reduction in academic staff over the next three or four years.

On undergraduate courses, the report says that because the proposed new national school curriculum will lead to a more broadly based, less specialized education for 16–18-year-olds, undergraduate programmes will have to be modified to include introductory material at the start. For this to work, undergraduate courses will have to be longer than the present three years, which would also prepare students better for PhD programmes, for which the usual three-year timescale is considered too short. The institute says: "Research students represent an important part of the manpower engaged in university and polytechnic research. Most meet the end of their studentships just when they become productive in their research, and that seriously reduces the contributions they are able to make."

Simon Hadlington

UK vice-chancellors propose part-taught PhD scheme

London

THE present debate about the system of awarding PhDs in Britain has been further fuelled by recommendations from a working group of the Committee of Vice-Chancellors and Principals (CVCP). The group's two chief proposals are that taught elements be included in doctoral programmes and that universities consider the introduction of the award of PhD 'with distinction' for theses with particular

merit. It estimates that about 10 per cent of PhDs would qualify for the grading. The proposals are based on the results of a survey of 68 university institutions and comparison between PhD systems in various countries. It was found that in most British universities there is provision for taught courses at PhD level, particularly in the natural sciences. In several departments and disciplines, taught elements are compulsory, with four institutions requiring all PhD students to attend taught courses. The group found that where there was only a small number of doctoral students, formal taught courses were rare. In such cases the group recommends that efforts be made to join with neighbouring institutions, making use of distance-learning technology.

The value of taught courses in research techniques and methodology, in the subject of the research itself and in related subjects, was clearly indicated by institutions that participated in the study. The working group comes out firmly against the notion of doctoral programmes involving a substantial taught component with only a limited research project, which a few institutions are said to be considering. In France, West Germany, the United States and Japan, importance is placed on course work and that in Japan and the United States students entering doctoral programmes are usually expected to have a master's degree. Furthermore, in West Germany, the United States, Japan and the Soviet Union, doctoral programmes include study and examination in a number of minor subjects.

On its recommendations to grade PhDs, the group says that the academic community and employers were concerned that the diversity of quality in doctoral theses was not reflected in the result. The group's proposals are being circulated to British universities for comments.

Simon Hadlington

TPA too costly for medical use

Washington

THE winds of fortune shifted last week for Genentech and its blood-clot-dissolving drug tissue plasminogen activator (TPA). The Ontario Medical Association of Canada recommended to its members that they should not administer the drug because its US\$2,000 per dose price is too high — ten times that of the other available drug, streptokinase.

Coupled with the US Health Care Financing Administration's decision two weeks ago not to reimburse hospitals for administering the drug to patients who qualify for government medical aid under Medicare, the news could mean a smaller market for TPA. US hospital officials estimate that continued use of TPA on Medicare patients will cost some hospitals up to \$700,000 per year. Most hospitals are expected to put pressure on cardiologists to use the cheaper streptokinase, even though the effectiveness of streptokinase drops unless it is used within an hour of when the heart attack begins.

The price of Genentech's stock has dropped 18 per cent over the past two weeks, despite encouraging data from a placebo-controlled mortality study of TPA, and a clinical trial comparing urokinase to TPA for pulmonary embolism (see *Nature* 332, 7; 1988).

Carol Ezzell

...and of course, you'd be entitled to write (PhD)² after your name...



merit. It estimates that about 10 per cent of PhDs would qualify for the grading.

The proposals are based on the results of a survey of 68 university institutions and comparison between PhD systems in various countries. It was found that in most British universities there is provision for taught courses at PhD level, particularly in the natural sciences. In several departments and disciplines, taught ele-

Power cuts fuel speculation

London

THE future of Britain's only large-scale fossil fuel experimental power station is in trouble now that the Central Electricity Generating Board (CEGB) has decided to pull out of the project. The decision brings to three the number of major research projects from which it has withdrawn this year (see *Nature* 332, 384; 1988).

The decision will leave British Coal, the other major sponsor, high and dry unless the directors of the Pressurised Fluidised Bed Combustor (PFBC) at Grimethorpe in Yorkshire can find another source of funds to meet the shortfall.

The plant was intended to demonstrate coal combustion with minimum release of sulphur and other pollutants. Instead of scrubbing them from the exhaust gases, sulphur is removed by the addition of dolomite limestone to the burning coal.

A report on the first £28-million phase of the project appeared in last month's *Journal of the Institute of Energy*. To install turbines in the 80-MW power station will require a further £12 million a year, which British Coal says it cannot afford by itself. The directors of the Grimethorpe project are this week due to meet representatives of the Paris-based International Energy Agency (IEA) to discuss alternative sources of funds.

Power stations built around the PFBC principles are claimed to be compact and versatile, burning various coal types at high pressure and relatively low temperature. Sulphur emission is said to be reduced by 90 per cent, while the stream of hot gases produced by the furnace can be used to drive a turbine, without the intermediary of a steam-raising plant, for an overall thermal efficiency of 41 per cent compared with the 38 per cent of a steam turbine system.

There has been substantial foreign interest in Grimethorpe, reflected in sponsorship from the US Department of Energy and the US Electric Power Research Institute. Plants based on the PFBC are under construction in the United States and in Sweden and are planned in Spain.

Ironically, the planned privatization of the British power generation industry may create a market for relatively small, efficient and environmentally friendly coal-fired stations. But the Grimethorpe announcement comes after a great deal of speculation about the future of the CEGB and its £162 million-a-year research commitment. The British governments white paper (policy document) on the privatization of the industry did not mention research (see *Nature* 332, 299; 1988).

Henry Gee

Australian vice-chancellors opposed to government reforms

Sydney

THE heads of Australia's universities registered their opposition this week to the government's proposed reforms of Australia's higher education system, saying they will lead to bureaucratic intervention by government in the detailed affairs of the universities.

The reforms, put forward by the new Minister for Employment, Education and Training, John Dawkins, require that each institution prepare an 'educational profile' detailing how national interests will be served. A proportion of the education budget, initially 1 per cent, will be awarded by the government according to the relative merit of these profiles. The institutions are also expected to obtain a much greater proportion of their research funds from industry and to concentrate their efforts in areas of special expertise.

Legislation now before parliament will establish a National Board of Employ-

ment Education and Training (NBEET) which will advise the minister directly. Under the board will come the Tertiary Education Council, responsible for recurrent funding for universities, and the Australian Research Council (ARC), responsible for research funding.

The Australian Vice Chancellor's Committee has now released a statement opposing this plan. The committee wants NBEET to be established as a statutory body, headed by a distinguished figure, so



Dawkins — serving national interests

that the universities will be "cushioned against the whims of government".

The vice-chancellors are concerned that the government is trying to centralize control over research funds without providing criteria on which to make decisions and without saying how national economic and social priorities are to be determined.

Even though it would provide the higher education sector with an advocate in the minister's court, a statutory NBEET would not necessarily be in the best interests of higher education, according to Professor Don Aitkin, interim chairman of ARC, as it would distance the minister from the institutions. Aitkin believes the minister should be allowed to govern.

The vice-chancellors also called for funding for higher education to be maintained at 1 per cent of gross domestic product (GDP). That implies annual growth of 3 per cent, well above the present figure of 0.99 per cent. Here Dawkins agrees, saying that, "what's inevitable is that the amount devoted to higher education in future will be at least 1 per cent (of GDP)". But Dawkins did not go so far as to promise to keep it there with government funds.

Charles Morgan

UK medical research in morale crisis

London

POOR pay and poor career prospects are seriously damaging the morale of medical researchers in Britain, according to a report from the House of Lords, published this week*. "It is not uncommon for a post-graduate research post to be advertised at a salary lower than would be paid a typist in many other fields, and moreover such posts are normally short-term contracts providing no career prospects whatever. Training grants are very low. These failings have to be remedied", the report from the House of Lords Select Committee on Science and Technology states.

But while poor pay and conditions are deterring many would-be medical researchers, the lack of adequately financed laboratories with sufficient technical support staff is an equally pressing problem. The committee calls for a greater proportion of grants to be given to long-term programmes rather than projects, and the allocation of £25 million towards the re-equipment of medical research laboratories.

The committee describes as unsatisfactory the funding of applied research in medicine and proposes the creation of a National Health Research Authority within the National Health Service which would take on a primary role in financing public health research and a part role in funding clinical research, but would not pay for basic research. Simon Hadlington

*Priorities in Medical Research. HMSO, London. £6.30.

UK information technology research programme criticized

London

BRITAIN'S national programme of research in advanced information technology has largely failed to live up to its original expectations, according to the independent scrutineers of public spending, the National Audit Office (NAO). The programme has been hampered by administrative weaknesses and shortages of suitably qualified personnel and will spend some £60 million more than initially envisaged, the NAO says in a report* published last week.

In an attempt to increase Britain's competitiveness in information technology in world markets, the government launched in 1983 the Alvey programme, with £200 million of public money, from the Department of Trade and Industry (DTI), the Science and Engineering Research Council and the Ministry of Defence, and £150 million from private industry. The administrative difficulties highlighted by the NAO have been acknowledged by the Alvey directorate. They arose chiefly because of the problems in coordinating three government departments with inadequate central administration. The directorate, based in the DTI, was severely understaffed so that highly skilled technical personnel had to spend time typing and filing.

The NAO points out that the directorate did not initially establish central management and financial information systems, making the provision of programme data more labour intensive and less comprehensive and reliable "than it should have been for a complex programme of this sort". The NAO acknowledges that in spite of the problems of administration, "which may have been partly caused by a lack of resources", the programme did draw up detailed strategies within 14 months and got 300 projects under way by March last year.

One of Alvey's chief objectives was to foster collaborations between industry, academics and research organizations. In this it succeeded to a largely commendable degree, the report says. But it qualifies its praise by noting that difficulties in establishing collaborative arrangements caused "significant delays" and withdrawals from the programme, and that project managers believed that the Alvey directorate could have helped more.

Despite the directorate's view that evidence of exploitable results emerging from nearly half of the 200 industrial projects by the middle of last year was "highly encouraging", the NAO says that the rate of exploitation was lower than originally foreseen. In particular, the software engineering programme, which had been

seen as having the greatest potential for rapid exploitation, had yielded no marketable products by mid-1987.

On the programme's finances, the NAO says that the Alvey board decided to spend some £35 million more on academic research and on programme administration and infrastructure than had been originally intended. So, total expenditure on Alvey research is likely to be around £60 million less than first thought.

Throughout the programme, skills shortages have hampered progress, particularly in the area of artificial intelligence, the report says. In more than half of the 42 projects studied by the NAO, manpower shortages were reported, resulting in delays, needs for extensions,

withdrawal of partners and the use of foreign experts. The original Alvey committee recognized the problems of skills shortages and recommended a specific role for the Alvey directorate to monitor manpower needs for the programme and to introduce action necessary to meet them. Such a role was not conferred.

Those closely associated with Alvey see the NAO report as concentrating too heavily on the problems encountered while failing to give sufficient credit for the part Alvey has played in promoting collaboration, particularly between companies whose researchers would probably not have otherwise had formal contact.

The NAO report will go before the House of Commons public accounts committee next month, when the agencies involved in the programme will be asked to comment.

Simon Hadlington

*Department of Trade and Industry: The Alvey Programme for Advanced Information Technology. HMSO, £5.20.

Indus Valley rock carvings under threat from road-building plans

London

THE rock-carvings of the Indus valley, a unique anthropological treasure spanning more than four millennia, are under grave threat of destruction according to Professor Karl Jettmar of Heidelberg University, addressing a conference on the religions of central Asia in London last week. Yet the sheer abundance of the carvings (more than 20,000, including more than 2,000 inscriptions, extending some 70 km along the Indus valley), make rescue impractical. The best hope, Jettmar said, is for archaeologists and anthropologists to make a detailed record of the carvings before they are lost.



Carving of stupas (interpreted as demons), traditional animal figure and inscription in Sanskrit alphabet.

The Indus Valley is the shortest natural route between Central Asia and what is now northern Pakistan. The carvings, preserved on account of the very dry climate, in the lee of the Karakoram mountains, reveal the shifting pattern of cultures and beliefs, making them what Jettmar called "the visiting card of history."

According to Jettmar, the first threat to the carvings was the decision of the Pakistan government to build a modern highway from Islamabad along the Indus valley to the Khungerab pass. Although the highway itself did relatively little damage, providing easier access when foreign archaeologists were allowed back into the area in 1979, the road seems to have inspired local people to build European-style houses, chopping up the nearest available rock for building blocks. Moreover, the new, fundamentalist mullahs, no longer prepared to accept the old folk-beliefs of the area, have convinced the locals that there was no supernatural danger in destroying the carvings.

An even greater threat, however, comes from the Tarbela reservoir, near Abbottabad. Although far downstream from the carvings, the reservoir is filling up fast with Indus silt and is likely to become useless in a few decades, possibly before it has paid for itself. And the planned relief reservoir at Basha will be only 30 km downstream from the richest area of carvings around Chilas village and will create a 40 km long lake drowning many of these carvings. The need to raise the level of the Karakoram highway by 200 m will also, Jettmar says, imperil many carvings not actually threatened with submersion.

Vera Rich

Creationism and common sense

SIR—While I take exception to the view held by Reginald T. Chelvam (*Nature* 331, 10; 1988), I will agree that “common sense” did not play a part in the Supreme Court’s decision. “Clear thinking” won the day. Chelvam’s “common sense” is that of a zealot. Common sense, to the general public, would accept that creation science is the valid competing scientific theory that rightly disproves or opposes darwinism. That is patently false and demonstrates the public unfamiliarity with either. Whereas darwinism is only a theory and lacks much evidence that it predicts should exist, there are opposing theories that are scientific and not creationist (that is, not based on the assumption of God’s existence). Those are the theories that should be taught in opposition to Darwin’s theory of evolution.

The cause of this dilemma is that the public perceives that because the Constitution separates church and state, such a schism must apply in science also, as if the Constitution were itself “holy scripture”. (This not to say that the Constitution has no merit; it is the last word when it comes to governmental processes, not physical ones. One cannot extrapolate its statements indiscriminately.) That extrapolation is false. Science and religion need not be an “either/or” proposition. Scientific inquiry cannot presume the existence of God as that would be self-contradictory, but such thought does not require denying the existence of God.

DAVID WAGGER

29 Park Drive, Apt 1,
Boston, Massachusetts 02215, USA

SIR—R. T. Chelvam takes issue with your leading article “Setback for creation science”, but entirely misses the point. Our current understanding of evolution, based on extensive experiment, threatens only the religions of those whose beliefs tend towards idolatry of ancient texts. It is certainly true that many eminent scientists of the past and present believe in the same creator as the ‘creation scientists’; the disagreement comes in the evolution of these beliefs. Scientists, whether with or without a degree in science, are committed to understanding our Universe, and its past, by experiment. As this understanding increases, old notions evolve into more appropriate views. On the other hand, ‘creation scientists’ are committed to imposing their belief in biblical inerrancy (usually applied only to the first creation story in *Genesis*) on an unsuspecting public. By fraudulently depicting evolution as the converse of creation, and then appealing to the public’s sense of fair play, they have tried to get their religious views into school science teaching. They have

failed; that, as you pointed out in an earlier leading article (*Nature* 327, 643; 1987), is the significance of the US Supreme Court’s decision against Louisiana’s “Balanced Treatment Act”.

Religious citizens can take heart from the Supreme Court’s decision; our understanding that all life forms are descended from a single progenitor is no more the death knell of religion than was the destruction of the belief in a geocentric universe in the sixteenth century. Those who would legislate religious beliefs would do well to consider the thoughts of Thomas Paine. As he pointed out in *The Rights of Man* (1787), such legislation usually confuses the mortal who worships and the immortal who is worshipped, and often demeans the latter. “Were a bill brought into parliament entitled ‘An act to tolerate or grant liberty to the Almighty to receive the worship of a Jew or Turk’ or ‘to prohibit the Almighty from receiving it’, all the men would startle, and call it blasphemy.”

ROGER C. PRINCE

Stonybrook Apiary,
Pittstown, New Jersey 08867, USA

SIR—Chelvam¹ attributes to me, without citation, remarks on molecular homology and its bearing on darwinism that he believes give support to creationism. I am said to have “referred to molecular homology (a latter-day prop of darwinism) as ‘anti-knowledge’ generating ‘antitheory’” and to “slop” in the data of molecular homology which have been “massaged with evolutionary theory.” Leaving aside the source or accuracy of these quotations, their use in this context betrays a misunderstanding of the role of homology in evolutionary biology that cannot go unchallenged.

Common descent provides an *explanation* for homology, and this is the only sense in which Darwin appealed to homology: “On this same view of descent with modification, all the great facts of Morphology become intelligible, — whether we looked to the same pattern displayed in the homologous organs . . . of the different species of a class; or to the homologous parts constructed on the same pattern in each individual”². Or “On my theory, unity of type [homology] is explained by unity of descent”². As homology is “always an inference, never an observation”³, it is plainly wrong to see the inference of common descent as a prop whose removal could affect the status of evolutionary theory.

Molecular homology, between amino-acid or nucleotide sequences, differs from classical homology in that its observational base has become statistical, and some molecular biologists have been castigated⁴

for conflating the observation of statistically significant similarity with the inference of relation through common ancestry. That may betray “muddy thinking”⁴ but it is not “slop” and it has no bearing whatever on the truth of descent with modification.

Chelvam asserts that “we are drowning” in evidence against darwinism. He cites nothing beyond the remarks attributed to me. It seems possible that he confuses two theories under the name of darwinism, the general theory of common ancestry or descent with modification, and Darwin’s special theory of mechanism, natural selection. If he knows of evidence inconsistent with the general theory of common descent, he should tell us what it is. I know of none.

COLIN PATTERSON

Department of Palaeontology,
British Museum (Natural History),
London SW7 5BD, UK

1. *Nature* 331, 10 (1988).
2. Darwin, C.R. *On the Origin of Species*, 456; 206 (John Murray, London, 1859).
3. Van Valen, L.M. *Nature* 305, 664 (1983).
4. Reeck, G.R. et al. *Cell* 50, 667 (1987).

Cup that cheers

SIR—Although a current study¹ shows that the aluminium level in infused tea is not after all as high as had earlier been reported², occasional reports of negative effects of tea-drinking^{3,4} must not be generalized as in most cases follow-up clinical studies are lacking. Nutritional and therapeutic values of tea are well documented⁵, and an international seminar in China recently stressed the importance of tea in improving human health. The beneficial effects of tea are mostly due to interactions between a number of compounds of tea infusions⁶, and a single compound may not be of consequence. For example, clinical studies show that the complex tea pigments are effective against cardiac diseases such as arrhythmia and atherosclerosis; black tea brew provided with 1–2 parts per million fluoride increases fluoride uptake⁸, and may thus help in preventing dental caries⁹. Despite some negative findings of a speculative nature, tea continues to cheer people the world over!

B. BANERJEE

Tea Research Association,
Nagrakata,
West Bengal 735225, India

1. Fairweather-Tait, S.J., Moore, G.R. & Fatemi, S.E.J. *Nature* 330, 213 (1987).
2. Corriat, A.M. & Gillard, R.D. *Nature* 321, 570 (1986).
3. Morton, J. Q. *J. Crude Drug Res.* 12, 1829 (1972).
4. Kapadia, G. *Carcinogenesis and Mutagenesis in Environment* Vol 2. (CRC, Baton Rouge, 1983).
5. Das, D.N., Ghosh, J.J., Bhattacharyya, K.C. & Guha, B.C. *Indian J. appl. Chem.* 28, 15 (1968).
6. Stagg, G.V. & Millin, D.J. *J. Sci. Food Agric.* 26, 1439 (1975).
7. Xia Shunying, *Int. Symp. Tea Quality Hum. Hlth.* 134 (1987).
8. Karunanayake, E.H., Mahadeva, K., Weerakoon, S.N. & Wickremasingha, R.L. *Tea Q.* 43, 111 (1972).
9. Yu Xingfu *Int. Symp. Tea Quality Hum. Hlth.* 157 (1987).

Licence to slang Copenhagen?

The passage of six decades has not ensured that the standard interpretation of quantum mechanics is generally accepted. If anything, discontents are growing stronger.

LESS than a year after the celebration of the centenary of Erwin Schrödinger's birth, the drum-beat of subdued discontent with the standard interpretation of quantum mechanics has become almost deafening. Not that serious interest in the meaning of quantum mechanics is all that new. In one sense, the argument about Bohr's interpretation has not ceased since the mid-1920s. One theme of the past few years, for example, has been the attempt to put the Einstein-Podolsky-Rosen paradox to the test of experiment, as in Aspect's experiment. But the novelty in more recent articles on the meaning of quantum mechanics is that people write as if they now have a licence to be puckish about Bohr and all his works.

There are several related but distinct arguments to be resolved, of which the chief is the physical nature of a quantum system, say an electron or a photon. Bohr's opinion, now called the Copenhagen view, is that such an entity is at all times *either* a wave *or* a particle, and that the doubt is resolved only when the entity interacts with some other system, as when some measurement is made of it. The contrary opinion, due to de Broglie, is that a quantum entity is at all times *both* a wave *and* a particle; the wave (the 'pilot wave') guides the particle in a manner made explicit by the version of quantum mechanics originally founded by David Bohm, but since marvellously elaborated.

Subsidiary issues are also important. All sides agree that the interpretation of Schrödinger's wave functions, ψ , as quantities whose squared amplitudes give the probability density that a particle will be at a particular place (if the arguments of ψ are space coordinates) was plucked like a rabbit from Bohr's hat. The strongest justification of the rule is that it works. But it remains an issue whether the rule applies to the understanding of a wave equation describing a single particle or whether it must be interpreted statistically, as the result emerging when identical measurements are made of a very large number of identical systems.

Linked with this is another teaser: what is the state of a quantum system after a measurement has removed the inherent doubt (as Copenhagen would have it) about its condition? The general opinion that a quick repetition of a measurement will yield the same result was neatly codified, in the 1940s, by von Neumann's argument that the effect of a measurement

(or, in general, any interaction with some other system) is to 'collapse' the state of a system onto a subset of all the states in which it might be found, technically by means of a projection operator acting within the space defined by all the orthogonal states of a system. But there are repeated arguments about the details of this process of collapse.

It goes without saying that these questions are embedded in a deep philosophical dispute. Copenhagen, sustained by the positivism derived from Mach, took the view that the circumstance that one cannot tell whether a quantum entity is a wave or a particle until it has interacted with a measuring system is a virtue, conforming as it does with the principle that physics should not be concerned with unmeasured or unmeasurable quantities. But this provokes the discontent of those who, like Einstein towards the end of his life, hold that there need be no restraint on the questions asked in physics.

That is the solemn side of the controversy. A. Posiewnik and J. Pykacz from the University of Gdansk put the argument differently and more entertainingly (*Phys. Lett.* **128**, 5; 1988), using a simple allegory to epitomize the dispute between positivist Copenhagen and (they imply, unfairly) the rest of the world:

Let us imagine a flat surface ('Flatland') inhabited by two-dimensional creatures ('Flats') who are able to investigate only those parts of surfaces of three-dimensional objects that are tangent to their two-dimensional world. A coin tossed on their surface appears to Flats either as heads or as tails and this problem is known in the scientific community of Flats as 'heads-tails dualism'. The early attempts at building up an exclusively heads or tails theory of coins failed, but it was pointed out by the group of scientists from the town of Flopenhagen that, in any experiment, any single coin reveals either its heads *or* tails character. This observation was proclaimed as the ultimate knowledge about coins and it was strictly forbidden to ask if a particular coin is 'heads' or 'tails' until the experiment was completed. Other groups of Flatland scientists tried to argue that a coin is both heads *and* tails (and maybe even something more) but...

The piece inevitably continues with the suggestion that Flopenhagen, believing that its interpretation constituted "a complete ultimate limit of knowledge", somehow arranged that views contrary to its own would be given a poor hearing.

All of this is good knock-about stuff, but with a serious purpose. As theoreticians will, the authors put forward a *gedanken* experiment designed to distin-

guish between Bohr and de Broglie, but in passing they let slip a revealing remark that will suggest to some that the dichotomy between the opinions attributed to the two giants of the 1920s may be an illusion. For there is an intermediate position, that occupied by the late Richard Feynman's view of quantum mechanics (see p.588), which Posiewnik and Pykacz call the "neo-Copenhagen" view.

Feynman's way of doing quantum mechanics is plainly, on reflection, a delicious compromise between Bohr and de Broglie. There are amplitudes (wavefunctions, or fields) which depend for their interpretation on the Copenhagen rule that the square of the amplitude is a probability density. But fields evolve in time (or in proper time, allowing for relativity) in a curiously classical fashion. One must enumerate all possible paths between A and B (with clever parameterization, not usually as difficult as it sounds), weight them according to the classical action along that path, add up amplitudes (which, being in general complex numbers, may yield interference effects) and then calculate squared moduli (to find probability densities).

This recipe does not eliminate what the philosophers regard as the worm at the heart of Bohr's philosophy, the *ex cathedra* declaration that algebraically additive amplitudes must be squared to obtain probability densities. That remains a rabbit from a hat. And, once accepted, that point of view implies all of what are popularly regarded as the nonsenses in the Copenhagen opinion — not least the notion that, in the absence of information to the contrary, for example, single particles travelling between A and B must be supposed in some sense to traverse all accessible paths.

The Gdansk *gedanken* experiment is also good fun. Posiewnik and Pykacz would use femtosecond laser pulses, fast shutters (say Kerr cells) and delay lines to chop pulses representing single photons in half and to arrange that the two halves travel by separate paths to a point at which (after one half has been delayed) they can interfere with each other. Suitable phasing of shutters in each path should make it possible to tell which way each half-photon has travelled, which on the Copenhagen view (but not Feynman's) should rule out interference. It will be interesting to see how it all comes out.

John Maddox

Protein dynamics

Hard facts on soft centres

Peter Artymiuk

LAST YEAR, Doucet and Benoit published in *Nature* a pioneering paper¹ on the interpretation of the diffuse scattering maxima in the background of X-ray diffraction patterns from protein crystals. They showed that, as in crystals of small molecules, it is possible to analyse some aspects of this diffuse scattering in terms of correlated molecular displacements. Their work, and that of Phillips *et al.*², provides a tantalizing glimpse of the wealth of information concerning the dynamical behaviour of biological macromolecules that could, in principle, be extracted from this seemingly intractable source of data (see my recent News and Views article³). But it seemed that it might be some time before generally applicable methods could be derived which could be used to obtain information of direct biochemical interest from these data. On page 659 of this issue⁴, however, Caspar, Clarage, Salunke and Clarage present a fascinating study of diffuse X-ray scattering, in which they derive evidence concerning both inter- and intra-molecular correlated displacements in crystals of insulin.

It is now generally accepted that co-ordinated molecular motions and molecular flexibility are crucial to the function of biological macromolecules, and that a full appreciation of the dynamics of proteins and nucleic acids is essential to elucidate their properties. The results obtained by Caspar *et al.* can be compared directly with, and indeed complement,

the information that is available from spectroscopic methods, from molecular-dynamics simulations⁵, and from crystallographic temperature factors⁶.

A key step in their interpretation is the identification of the underlying distribution of the scattering with a commonly used crystallographic self-correlation function called the Patterson function. Information about the range and amplitude of coupled displacements within the structure is obtained by theoretical calculations of the diffuse scattering, modified by different range and amplitude parameters. These results are then compared with the experimentally observed patterns.

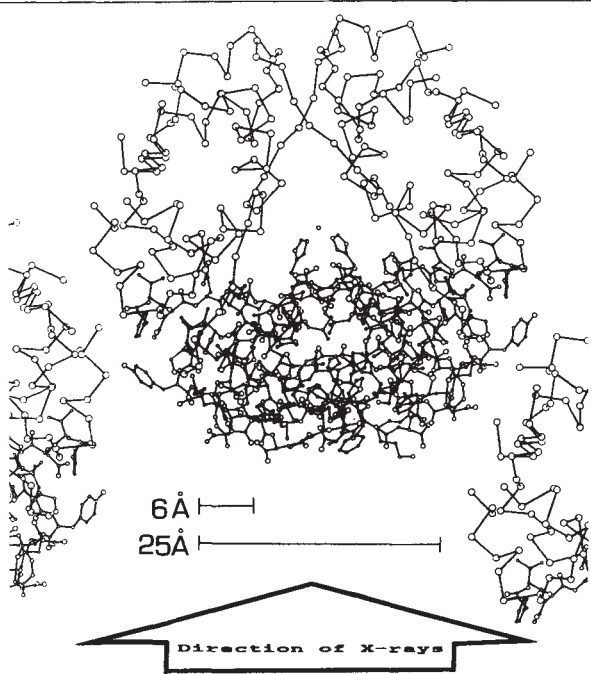
Caspar *et al.* identify two distinct types of coupled motion in insulin, each giving rise to a characteristic type of diffuse (or, as they call it, variational) scattering. First, after removing the effects of incoherent (Compton) scattering and the scattering resulting from disordered water, they show that the remaining general diffuse background arises from short-range correlated displacements (with r.m.s. amplitude 0.4 Å) of side-chains or parts of the protein background over a range of 4–8 Å. These are the kinds of motions that have had to be postulated to explain various processes in biological macromolecules, for example, to explain how oxygen can reach the haem group in myoglobin through a seemingly impenetrable mass of closely packed side-chain atoms⁷. Caspar *et al.* show that the second aspect of the diffuse scattering analysed,

which consists of sharper haloes around the ordinary Bragg diffraction peaks, arises from displacements (coupled over a range of 20–30 Å) of insulin molecules relative to one another within the crystal. This result is very exciting, as in favourable cases such studies could be extended to correlated motion of subunits in allosteric proteins such as haemoglobin⁷ or aspartate transcarbamoylase⁸, where large-scale movements are known to occur between R and T states. In their pioneering study⁹, Boylan and Phillips analysed diffuse scattering from tropomyosin crystals in terms of large-scale anisotropic motions and attempted to relate their results to the function of the protein in muscle.

There are several ambiguities inherent in the interpretation of the diffuse scattering data. The most important of these is the complete absence of information concerning timescales of the 'motions' detected by crystallographic methods in general³: this is why the neutral word 'displacements' is usually used, so that the possibility of static disorder can be included as well. However, much of the information obtained from crystallographic temperature factors is consistent with interpretation in terms of atomic motions¹⁰, and this could well also be true of much diffuse scattering data⁹. Another ambiguity in the diffuse-scattering study⁴ of Caspar *et al.* is that, whereas it is possible to say that there are coupled motions of atoms separated by 4–6 Å, and of insulin molecules 20–30 Å apart, it is not possible to say precisely which parts of the structure are involved. At present Caspar *et al.* base their analyses on one still photograph, but they have already begun to extend their studies into three dimensions by the use of area-detector data⁴. This, together with the evidence from the crystallographic temperature factors, should help resolve these questions.

It is clear that the elegant methods derived by Caspar *et al.* should lead to much more hard information becoming available about concerted displacements of protein subunits — and of liquid-like movements of side-chains and main-chain segments — both on the surface and in the centre of the protein. □

An insulin molecule in the crystal¹¹. Three dimers form a hexamer around a crystal 3-fold axis. One dimer is shown in atomic detail, the other two as backbone traces. The 6- and 25-Å scale bars represent respectively the average distances of the short-range and of the long-range coupled motions detected by Caspar *et al.*⁴. Large arrow, the approximate direction of the X-ray beam in their experiment. In the crystal, each hexamer is completely surrounded by others: parts of just two of these are shown at the lower edges of the figure.



1. Doucet, J. & Benoit, J.P. *Nature* **325**, 643–646 (1987).
2. Phillips, G.N. *et al. Biophys. J.* **32**, 485–502 (1980).
3. Artymiuk, P. *Nature* **325**, 575–576 (1987).
4. Caspar, D.L.D., Clarage, J., Salunke, D.M. & Clarage, M. *Nature* **332**, 659–662 (1988).
5. Karplus, M. & McCammon, J.A. *A. Rev. Biochem.* **53**, 263–300 (1983).
6. Artymiuk, P. *et al. Nature* **280**, 563–568 (1979).
7. Perutz, M.F. *Scient. Am.* **239**(6), 68–86 (1978).
8. Krause, K.L. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 726–739 (1970).
9. Boylan, D. & Phillips, G.N. *Biophys. J.* **49**, 76–78 (1986).
10. Petsko, G.N. & Ringe, D.A. *A. Rev. Biophys.* **13**, 331–371 (1984).
11. Dodson, E.J. *et al. Can. J. Biochem.* **57**, 1200–1203 (1987).

Peter Artymiuk is in the Department of Biochemistry, University of Sheffield, Sheffield S10 2TN, UK.

Evolutionary biology

Sexual selection by female choice in monogamous birds

Rory Howlett

DARWIN'S proposal¹ that preferences expressed by female birds when choosing mates could drive the evolution of male ornamentalities like the peacock's tail caused immediate uproar. Critics such as Wallace accused him of invoking an anthropomorphic aesthetic sense on the part of females: not only was the proposed existence of this sense unexplained, it also ran counter to the critics' panselectionist philosophies. This suspicion remained deeply ingrained until the early 1930s, when Fisher suggested² how female choice and male ornamentation could co-evolve through a process of positive feedback — the famous 'runaway process'. But male ornamentation in monogamous species, where the selection generated by female choice is expected to be weak, is an outstanding problem. On page 640 of this issue³, Anders Pape Møller presents experimental evidence that female choice selects for male sexual ornamentation in a monogamous swallow.

Fisher believed that two fundamental conditions must be fulfilled if Darwin's theory of sexual selection by female choice is to account for the evolution of male sexual ornamentation: the existence of sexual preference in females; and the prevalence of life-history patterns that confer a reproductive advantage on sexually preferred individuals. Taking the second condition, the intensity of sexual selection in polygamous birds is potentially high as a few males can inseminate many females; that is, the variance in the mating success of males is large. Indeed, there is strong experimental evidence that selection by female choice has selected for male sexual adornments in polygamous birds. In a classic study of the widowbird, *Euplectes prognus*, Andersson showed⁴ that males with experimentally elongated tails have a greater mating success than those with shortened tails. Andersson excluded

the possible confounding effects of territory quality by randomizing the experimental birds across territories, and his behavioural observations suggest that direct competition between males is not responsible for the observed results either.

Thus, it seems that sexual selection by female choice can indeed give rise to the evolution of male sexual ornaments in polygamous birds. But what of monogamous birds? Here, there should be little variance in the number of mates that males obtain, so the power of sexual selection is expected to be weak. But many monogamous species seem to have undergone extreme sexual selection with males that are as splendid as their polygamous counterparts⁵. This paradox was evident to Darwin who, demonstrating his usual clarity of thought, proposed what Fisher described as an 'exceedingly subtle' explanation. The essence of his idea is that in monogamous species preferred males tend to breed earlier in the season, and if earlier breeding is correlated with increased reproductive success, preferred males would benefit by leaving more offspring. Fisher thought that the demonstration of such a correlation, whatever its causes, would not be easy. In fact, such a correlation has now been established⁶ for many birds but few studies have demonstrated its role in sexual selection.

One exception is a long-term study⁶ by O'Donald of the monogamous arctic skua, *Stercorarius parasiticus* (see figure). Melanic individuals of this species have a general reproductive advantage over pale ones in part because of an advantage in sexual selection. The fledging success of pairs of arctic skuas declines as the breeding season proceeds, and it seems that darker males take less time to find a mate than pale ones, breed earlier, and benefit as a result. This difference is evident both for experienced males who have bred in

previous years and for new males breeding for the first time. The fact that male experience seems not to affect the time to find a mate suggests that female choice rather than male competition is responsible for the earlier breeding times of melanics.

This conclusion is supported by O'Donald's success in fitting data on breeding dates of males with theoretical models which assume that preferential matings precede random matings in the breeding season. These models predict that there will be linkage disequilibrium between the gene for melanism and the gene for preference. Thus, choosy females will tend to be melanic and assortative mating will result — a prediction fulfilled by the data. Although his evidence for sexual selection by female choice in the arctic skua is indirect, relying on demographic studies rather than experiment or direct observation of courtship behaviour, O'Donald concludes that sexual selection by female choice can operate in monogamous birds just as predicted by Darwin.

But the variation in the arctic skua is atypical in that it is discontinuous and under simple genetic control: most sexually selected traits, such as the peacock's tail, are quantitative and presumably under complex genetic control. Can conclusions from O'Donald's study and others like it be generalized to situations such as these?

The answer seems to be yes. In this issue³, Møller reports the results of his study of sexual selection and variation in reproductive success in a monogamous swallow (*Hirundo rustica*). In this species, the outermost tail feathers of the males are longer than those of the females, and appear to have evolved by sexual selection. Møller has borrowed Andersson's technique⁴ of experimentally manipulating the length of the birds' tails by glueing on extra lengths of feather or by cutting the feathers to investigate directly the effect of tail length on reproductive success.

As was the case in Andersson's widowbirds, male swallows with experimentally lengthened tails have a higher reproductive success than males with reduced tail lengths. This increased reproductive success results from two causes. First, males with longer tails tend to obtain mates more easily and thus mate earlier in



Chicks of the arctic skua (*Stercorarius parasiticus*). This species is monogamous and is polymorphic for three colour patterns: pale (left); intermediate (centre); and melanic (right). The polymorphism seems to be controlled by a single gene⁶: dark shows incomplete dominance over pale, with most intermediate birds being heterozygous.

the season, allowing them time to raise an extra brood. Second, the extra-pair-bond copulation rates of males with elongated tails are significantly higher than those of other males. As in the previous study, observations of the courtship behaviour of the swallows point strongly to female choice rather than male competition as being responsible for the increased reproductive success of males with longer tails. Møller's results not only help vindicate Darwin's theory, they also represent the first direct demonstration of an advantage

for males with extreme traits in a monogamous species, presumably as a result of a female mating preference. □

1. Darwin, C. *The Descent of Man and Selection in Relation to Sex* (Murray, London, 1871).
2. Fisher, R. A. *The Genetical Theory of Natural Selection* (Clarendon, Oxford, 1930).
3. Møller, A. *Nature* **332**, 640–642 (1988).
4. Andersson, M. B. *Nature* **299**, 818–820 (1982).
5. Bradbury, J. W. & Andersson, M. B. eds *Dahlem Workshop on Sexual Selection: Testing the Alternatives* (Wiley, Berlin, 1987).
6. O'Donald, P. in *Avian Genetics* (eds Cooke, F. & Buckley, P. A.) 433–450 (Academic, London, 1987).

Rory Howlett is an assistant editor of *Nature*.

Climatology

Solar cycles and the atmosphere

Marvin A. Geller

THERE is no more controversial topic in meteorology than that of whether solar activity exerts a significant control on weather and climate. Karin Labitzke of the Free University of Berlin and Harry van Loon of the National Center for Atmospheric Research show in a provocative paper¹ published this week that effects of the 11-year solar cycle are discernible if atmospheric data are subdivided according to the phases of the 2–3-year cycles in equatorial wind patterns (the quasi-biennial oscillations, QBOs). In particular, they analyse atmospheric data from the Northern Hemisphere for the months of January and February since 1956. They divide the data into two portions: one in which the equatorial mean zonal winds at an altitude of about 20 km are westerly, and the other where they are easterly, termed the westerly and easterly phases of the QBO. They find that, although there is no clear correlation between atmospheric behaviour and an index of solar activity (the 10.7-cm radio flux) for the whole data set, there are strong positive and negative correlations in the subsets of the data during the individual phases of the QBO (Fig. 1).

This apparent response of the atmos-

phere to solar activity is fairly large in the lower stratosphere (10–25 km altitude) and appears to extend downwards well into the troposphere (below 10 km). Moreover, maps of these correlations display a planetary-wave-like distribution (Fig. 2) that reverses sign with the changing phase of the QBO. Also the correlations at high latitudes are of opposite sign from those at lower latitudes. These signatures suggest that interactions of planetary waves with the mean zonal atmospheric state may be involved. Previous observational analyses² indicate that there is significant QBO modulation of the mean zonal wind and the planetary waves in the extratropical stratosphere during winter in the Northern Hemisphere. Although no detailed explanation for this exists, it is thought that whether the direction of the equatorial winds is easterly or westerly affects the nature of the stratospheric waveguide for the planetary waves and thus can affect stratospheric planetary-wave structure and their interaction with the mean flow.

There are two important issues that arise from Labitzke and van Loon's paper. First, do the statistics indicate a clear association between solar activity and

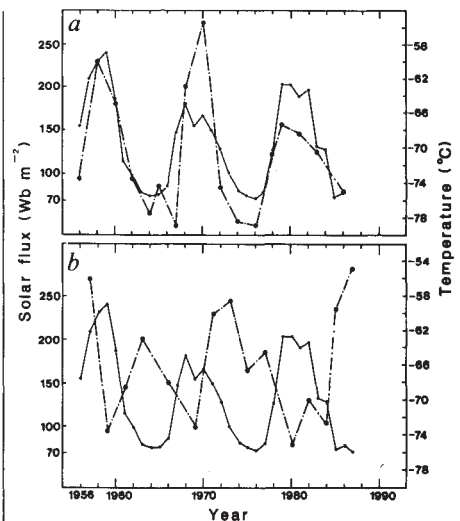


Fig. 1 Comparisons of the 10.7-cm solar flux (broken lines) and the atmospheric temperature 23 km above the North Pole (solid lines) for the westerly (a) and easterly (b) phases of the QBO. All data are averages for January and February of each year. The correlation (a) and anticorrelation (b) are evident; no such correlation is apparent for the complete data set. (From ref. 1).

atmospheric structure? The authors apply reasonable statistical tests to show that the determined correlations are almost certainly not a chance occurrence. Note, however, that their data set covers only three 11-year solar cycles. There have been previous examples³ of published correlations between some geophysical variable and solar activity that break down after only a few solar cycles. At one time there seemed to be strong evidence that the depth of Lake Victoria correlated with solar activity, but the statistical association turned out to be transient.

Second, is there a credible mechanism to explain the new correlations? Several models^{4–6} of the response of the atmosphere to the cyclic variation of solar ultraviolet irradiance indicate a much smaller variation in temperature than that found by Labitzke and van Loon. But these studies use only two-dimensional models (altitude and latitude). One calculation⁶ that combines a two-dimensional model with a planetary-wave propagation model shows much smaller planetary-wave responses than are found by Labitzke and van Loon. These results do not necessarily indicate that there is no viable mechanism for Labitzke and van Loon's correlation, however. Ramanathan and co-workers, using a general circulation model of the atmosphere, find⁷, in high-latitude, lower-stratospheric regions, a much larger response to rather small changes in the treatment of radiative transfer than might have been expected. After detailed examination of these results and comparisons with other models, they conclude that the effect of feedback between dynamical and radiative processes in the model is much larger than the direct

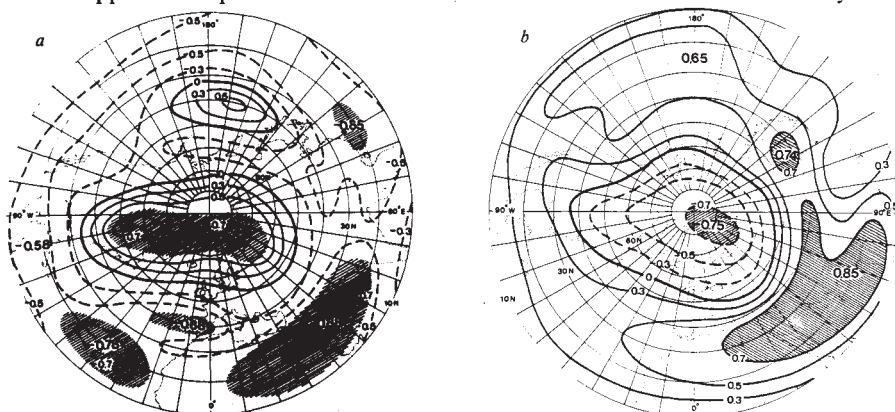


Fig. 2 Contour plots of the correlation coefficients of atmospheric temperature at 20 km and solar activity for the westerly (a) and easterly (b) phases of the QBO. The contours (solid for positive values, broken for negative) show a large-scale planetary-wave type pattern, with the values having opposite sign in the two phases. (From ref. 1.)

radiative effects. It seems that radiative changes can alter the meridional (Equator-to-pole) temperature gradient, which in turn changes the polar vortex. This alters the planetary-wave propagation, which further changes the vortex, so that the process iterates. If the atmosphere displays the same sort of sensitivity as does this model, the findings of Ramanathan *et al.* indicate a viable mechanism for Labitzke and van Loon's results.

There has been a resurgence of interest in the effects of solar activity on the atmosphere, particularly the middle atmosphere, during the past two years, probably because of the Middle Atmosphere Program and recent satellite observations of the solar ultraviolet irradiance and its variations. This resurgence will probably continue with the forthcoming Solar Terrestrial Energy Program. Labitzke's work, both by herself and with van Loon, are leading examples of these recent studies. These have led T. Dunkerton to examine very similar data⁸, particularly looking for the height dependence of the correlations. For example, he finds that during the westerly phase of the

QBO, the correlation of polar temperatures with solar activity found at 20 km is not found at 23 km. Dunkerton suggests that the vertical structure of these effects needs to be examined more closely. M.L. Chanin⁹ finds indications of more significant solar-activity control of middle atmosphere temperatures, as measured by the Rayleigh lidar technique, than are predicted by two-dimensional models. No doubt such analyses of atmospheric response to solar activity will motivate new directions in atmospheric modelling. □

1. Labitzke, K. & van Loon, H. *J. atmos. terr. Phys.* **50**, 197-206 (1988).
2. Holton, J.R. & Tan, H.-C. *J. atmos. Sci.* **37**, 2200 (1980).
3. Pittcock, A.B. *Rev. Geophys. Space Phys.* **16**, 400 (1978).
4. Brasseur, G. & Simon, P.C. *J. geophys. Res.* **86**, 7343 (1981).
5. Garcia, R., Solomon, S., Roble, R.G. & Rusch, D.W. *Planet. Space Sci.* **32**, 411-423 (1984).
6. Callis, L.B., Alpert, J.C. & Geller, M.A. *J. geophys. Res.* **90**, 2273-2282 (1985).
7. Ramanathan, V., Pitcher, E.J., Malone, R.C. & Blackmon, M.L. *J. atmos. Sci.* **40**, 605-630 (1983).
8. Dunkerton, T.J., Delisi, D.P. & Baldwin, M.P. *Geophys. Res. Lett.* **15**, 136-139 (1988).
9. Chanin, M.L., Smires, N. & Hauchecorne, A. *J. geophys. Res.* **92**, 10933-10941 (1987).

Marvin A. Geller is at the Laboratory for Atmospheres, NASA/Goddard Space Flight Center, Greenbelt, Maryland 20771, USA.

Oceanography

Heat flux in the Antarctic

Roland A. de Szoeke

In simple terms, the Earth acquires heat by radiation from the Sun in low, tropical latitudes and radiates it back to space in high, polar latitudes. The circuit is completed by transporting heat meridionally (north-south) through the dual-fluid envelope of the Earth: air and water. This is intimately connected with the variation of weather and climate over the planet and over seasonal to geological timescales. An especially interesting region is the Southern Ocean, the only place where there is a circumpolar belt of water uninterrupted by land masses so that a massive eastward current, carrying about $130 \times 10^6 \text{ m}^3 \text{ s}^{-1}$ of water¹, can exist. Heat must be transported across this current and this is apparently accomplished by an 'eddy flux' process for which Keffer and Holloway propose, on page 624 of this issue², a parameterization. The amount of heat transported can be estimated using this, from measurements by satellite-borne instruments of low-frequency (periods of several days and longer) variability of sea-surface height.

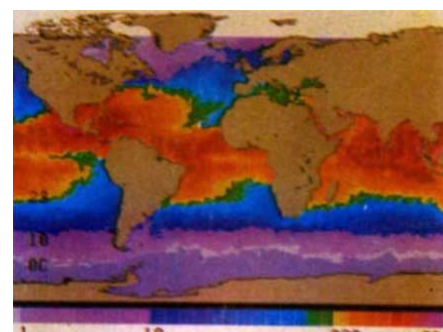
The atmosphere and oceans contribute roughly equally to the transport of heat from equator to poles, though the mechanisms differ. In low latitudes in the atmosphere (from the Equator to about 30° in either hemisphere), most heat is carried by the Hadley cell, a structure consisting of warmed air rising and flowing poleward and being replaced by cooler air flowing

equatorward at lower levels. In mid-latitudes (30°-60°) the poleward heat transport is effected by transient eddy fluxes — the weather systems in which poleward flow is correlated with warmer air, equatorward flow with cooler air. Although there are important zonal (east-west) variations in the atmosphere, associated with land-ocean contrasts, this simple, zonally uniform picture of the atmospheric heat cycle is largely adequate³.

Oceanic heat flux shows a very different pattern, because the oceans are interrupted by roughly meridional land masses forming two ocean basins in the Northern Hemisphere (we treat the Arctic Ocean as part of the Atlantic, as its connection to the Pacific is very weak), and three in the Southern Hemisphere, connected in a high-latitude circumpolar band in the Southern Ocean, and across the Indonesian archipelago between the Pacific and Indian. The lateral confinement of the ocean basins, except in the Southern Ocean, means that the average circulation in them consists of a series of asymmetric gyres, intensified into strong narrow currents like the Gulf Stream and the Kuroshio on the western sides of the basins. In the subtropical Atlantic, for example, the Gulf Stream transports northwards through the Florida Straits on average $30 \times 10^6 \text{ m}^3 \text{ s}^{-1}$ of water that is 8 °C warmer than the water returning south —

a northward transport of heat at the rate of $1.1 \times 10^{15} \text{ W}$ (ref. 4). Similar estimates for the North Pacific are indeterminate, perhaps because data have not been collected at consistent places. Still, it seems that the gyral transport of heat is much less than in the North Atlantic. In the South Atlantic, the oceanographic data are enigmatic — equatorward heat transport⁵. In the western equatorial Pacific, there is transport of warm water through the deep passages around the Indonesian islands, taking heat into the Indian Ocean and on to the South Atlantic in an intense current that rounds southern Africa⁶ (see figure).

Thus, the oceans gain heat at the surface in the tropical bands. In the Pacific most of this is transported south into the Southern Ocean, or leaks into the Indian Ocean where it joins the heat gained in the Indian tropical band and thence splits into two paths. One part flows into the Southern Ocean, where heat is lost to the atmosphere at high latitudes, the rest is carried



Satellite image (NOAA/Richard Legeckis) of sea-surface temperatures (colour scale in °C) during an Atlantic El Niño event in 1984.

round southern Africa in the intense Agulhas Current into the south Atlantic where it is carried equatorward to join the Atlantic's share of tropical heat gain and thence poleward through the North Atlantic and into the Arctic Ocean where it is lost to the atmosphere. The Atlantic is the only channel for oceanic transport of heat to the Arctic because the North Pacific is virtually blocked off at its northern extremity. This accounts for the anomalous equatorward transport of heat through the South Atlantic.

The heat transported across the Antarctic Circumpolar Current in the Southern Ocean to be lost to the atmosphere in high southern latitudes is estimated⁷ to be $0.30 \times 10^{15} \text{ W}$. Apparently, this is carried not by a vertically overturning cell as in the tropical atmosphere, nor by a horizontal gyre as in the laterally confined parts of the ocean basins, but is carried in the form of an eddy flux of heat as in the mid-latitude belts of the atmosphere. Keffer and Holloway propose in this issue² that this flux be monitored from satellite measurements of low-frequency variations of sea-level height. Their method depends on a variant of mixing-length

theory, commonly used in fluid mechanics and engineering to estimate diffusion by turbulence in pipes and channels. The turbulent transport of a quantity, for example, heat, is parameterized as a mixing coefficient times the average gradient of, in the case of heat, temperature. The theory provides an estimate of the mixing coefficient in terms of certain gross statistics of the turbulence — the root-mean-square (r.m.s.) velocity and the size (mixing length) of the eddies. The difference in the oceanic context lies in the large scales (10–100 km) of the turbulent eddies and the dynamics, involving the Coriolis force associated with Earth's rotation, that sustains them. The mixing coefficient is proportional to the r.m.s. sea-surface height, which serves for pressure in the ocean and hence for velocity, and which can be measured globally to sufficient accuracy (several centimetres) with radar altimeters on satellites. Just such an instrument was carried by the short-lived Seasat mission⁸ in 1978, the data from which Keffer and Holloway use to test their theory. Other altimeters are now in space or planned.

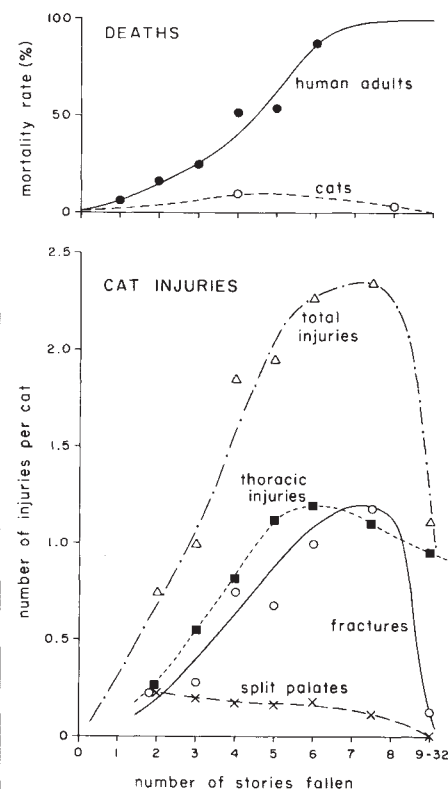
Keffer and Holloway's model is not flawless. The basic theory is worked out for geostrophic turbulence in a homogeneous layer of fluid on a curved, rotating Earth. The real ocean is certainly not homogeneous, and density stratification influences the vertical structure of large-scale eddies, which are invariably more intense near the ocean surface. The authors allow for this by arbitrarily halving the theoretical mixing coefficients. On the other hand, the Seasat data they used spanned only a 28-day period, which does not cover the range of eddy timescales (10–100 days) thought to be responsible

for turbulent heat flux. Keffer and Holloway compensate for this by doubling their mixing coefficient estimate. Although these corrections cancel, their arbitrariness introduces uncertainty. Also, the theory does not allow for the dynamical effects of buoyancy forces associated with temperature variations in eddies but instead treats temperature as a passive quantity. The large-scale temperature gradients associated with the Antarctic Circumpolar Current can produce eddies of the requisite time and space scales by 'baroclinic' instability⁹, which is absent from Keffer and Holloway's model.

Still, the new calculations give estimates for eddy heat flux across the Southern Ocean of the same magnitude as the heat lost by high-latitude waters. Although the theory is not watertight, it shows how satellite-borne instruments can yield ocean-wide data useful for computing the parameters, like eddy heat flux, of vital geophysical importance. *In situ* field experiments will test the details of theories and parameterizations such as Keffer and Holloway's, so that reliable methods of remote sensing with satellites can be developed. □

1. Fandry, C. & Pillsbury, R.D. *J. phys. Oceanogr.* **9**, 449–455 (1979).
2. Keffer, T. & Holloway, G. *Nature* **332**, 624–626 (1988).
3. Palmén, E. & Newton, C.W. *Atmospheric Circulation Systems* (Academic, New York, 1969).
4. Bryden, H.L. & Hall, M.M. *Science* **207**, 884–886 (1980).
5. Bennett, A.F. *J. phys. Oceanogr.* **8**, 785–798 (1978).
6. Gordon, A.L. *J. geophys. Res.* **91**, 5037–5046 (1986).
7. Gordon, A.L. & Owens, W.B. *Rev. Geophys.* **25**, 227–233 (1987).
8. Cheney, R.E., Marsh, J.G. & Beckley, B.D. *J. geophys. Res.* **88**, 4343–4354 (1983).
9. Pedlosky, J. *Geophysical Fluid Dynamics* (Springer, New York, 1979).

Roland A. de Szoeke is in the College of Oceanography, Oregon State University, Corvallis, Oregon 97331, USA.



Mortality rates for falling adult humans and cats (above), and number of total injuries and various types of injury per falling cat (below), as a function of number of stories fallen. (Based on the work by Waring and Demling and by Whitney and Mehlhoff.)

R.H. *Ann. emerg. Med.* **15**, 1088–1093; 1986). As illustrated in the figure, higher falls are increasingly lethal for humans, and few adults survive falls of more than six stories onto concrete. The principal causes of death are head injuries and haemorrhage from visceral injuries. Although forelimb fractures are slightly commoner than hindlimb fractures in falling cats, falling adult humans most often break their legs, and falling children their arms (Smith, M.D. *et al. J. Trauma* **15**, 987–991; 1975).

Straightforward theory relates injuries from falls to three sets of variables, as shown by Warner and Demling. First, the height of the fall determines the impact velocity. Second, the softness of the surface of impact affects the stopping distance and hence the impact force. Those people surviving falls from aeroplanes have landed on mud or snow, not concrete. And third, at least five properties of the falling body itself are relevant: its mass, which determines impact force (F) and energy; its cross-sectional area A , determining frictional drag during fall and also stress on impact (F/A); cross-sectional areas of bone, determining bone strength; cushioning of vital parts by fat and other soft tissue; and dissipation of impact forces through flexing of muscles and use of joints.

These theoretical considerations pro-

Animal behaviour

Why cats have nine lives

Jared M. Diamond

THE famous adage that cats have nine lives stems in part from their ability to survive falls lethal to most people. This phenomenon has not received the scientific attention that it deserves. Filling this lacuna, a new study by W.O. Whitney and C.J. Mehlhoff (*J. Am. Vet. Med. Assoc.* **191**, 1399–1403; 1987) applies principles of anatomy, physics and evolutionary biology to falling cats.

The authors were veterinarians at an animal hospital in New York City, where skyscrapers, open windows and paved ground combined to generate a database of 132 cats injured by falls of 2 or more stories, with a maximum of 32 stories and a mean of 5.5 ± 0.3 (s.e.m.) (1 storey = 15 feet). Most victims landed on concrete after a free-fall. Omitting 17 cats that were

euthanized by owners unable to afford treatment, 90 per cent of the cats (104 of 115) survived, whereas 11 died (mainly because of thoracic injuries and shock). The most remarkable feature of the results (see figure) is that incidence both of injuries and of mortality peaked for falls of around seven stories and decreased for falls from greater heights. For instance, the cat that free-fell 32 stories onto concrete was released after 2 days of observation in the hospital, having suffered nothing worse than a chipped tooth and mild pneumothorax.

Falling adult humans differ from falling cats in their much higher mortality rate, monotonic mortality/height relation, different causes of death, and different sublethal injuries (Warner, K.G. & Demling,

vide several reasons why cats survive falls that kill adult humans. First, because mass increases as the cube but surface area as the square of linear dimensions, falling large animals are in general more injury-prone than small ones, as they suffer greater impact stress, their bones experience greater stress, and they reach higher terminal velocities in free-fall because of a less favourable area/mass ratio. Even a small drop breaks an elephant's leg, but falling mice reach terminal velocity in the atmosphere much sooner and at a much lower value than do falling elephants.

Second, falling cats have a superb vestibular system and make gyroscopic turns such that all four feet are soon pointing downwards, regardless of the cat's orientation at the start of the fall. Hence cats dissipate the impact force over all four limbs. Falling human adults tend to tumble uncontrollably but land most often on two feet, next most often on their heads. Falling babies, because their relatively large heads shift their centre of gravity towards the head, tend to land head-first with arms reflexly extended to break the fall. These facts contribute not only to the lower mortality of falling cats but also to the tendencies of falling babies, adults and cats to broken arms, broken legs and breaks distributed over all four limbs, respectively.

Third, a cat falling in the atmosphere reaches a terminal velocity of about 60 m.p.h. (compared with 120 m.p.h. for adult humans) after only about 100 feet. As long as it experiences acceleration, the cat probably extends its limbs reflexly, but on reaching terminal velocity it may relax and extend the limbs more horizontally in flying-squirrel fashion, thus not only reducing the velocity of fall but also absorbing the impact over a greater area of its body. This may explain the paradoxical decrease of mortality and injury in cats that fall more than 100 feet.

Finally, cats that land with their limbs flexed dissipate much of the impact force through soft tissue. Parachutists are trained to dissipate impact forces by landing with knees and hips flexed, then rolling.

Evidently, falling cats have some advantages shared with any small animal of similar mass and shape but also have unique advantages of their own, notably their gyroscopic righting reflex and their limb flexing on landing. Small dogs that fall from buildings are prone to more serious injuries than cats. The cat-specific advantages have undoubtedly evolved through natural selection: most felid but few canid species are arboreal, so that millions of years of springing or falling from trees have favoured those felids with the best vestibular systems. Thus, the nine lives of cats are a product of their evolutionary history. □

Jared M. Diamond is Professor of Physiology at the University of California Medical School, Los Angeles, California 90024, USA.

Indian maize in the twelfth century BC



GODDESSES and gods in sculptured soapstone friezes in Hoysala temples of the twelfth and thirteenth centuries BC near Mysore, India hold in their hands representations of maize ears. There are more than 63 of these large ears at Somnathpur, and maize is represented at three other temples I have visited.

In the Hoysala tradition, worshippers must have used maize as a golden coloured and a many-seeded fertility symbol in their religious rites. That the ears are modelled on maize is shown by the ear length-to-diameter ratio, the ear sizes in relation to parts of human figures, and the wide variation of anatomical detail in the carvings that all belong to maize: the ears have either parallel, highly tapered or bulging sides, their tips are pointed, and their axes may be straight or warped, depending on the moisture at the time of picking and the way maize dries.

Normally, the husks are removed from religious offerings of maize to show perfection. Only part of the husks have been removed on a few ears in the sculptures, showing three to seven rows of kernels. Occasionally, husks are left whole with an etched curl of silk signifying that the smoothed object, with the several shapes of ears inside, is maize.

Kernels generally have a rounded, rectilinear outline with a width-to-thickness ratio that is found in modern maize, though it is more typical of ancient maize. The kernels get smaller near the tips, and kernels are larger next to a gap in the row as the adjacent kernel tends to fill in the space. Kernels are arranged generally in pairs, in parallel rows that nest together on the cob, and they shift relative arrangements up and down the ear the way the pairs of modern maize kernels do. The most definitive evidence is that a few ears illustrate that the maize models had a jumbled or tessellate arrangement at the butt of the ear.

No other plant or object has the extensive intricacy and variation of highly segregated maize that could serve as a model for the sculptures. No other fruits have the same number and shape of the closely packed kernels that are arranged in parallel rows in the sculptures.

The existence of maize in twelfth-century India appears to contradict the widely-held view that no evidence exists for trans-oceanic contact between civilizations at this period.

Carl L. Johannessen

Carl L. Johannessen is Professor of Geography at the University of Oregon College of Arts and Sciences, Eugene, Oregon 97403-1218, USA.

Richard Phillips Feynman (1918–1988)

RICHARD FEYNMAN, who died on 15 February, was one of the greatest physicists since the Second World War and, I believe, the most original. He won the Nobel prize for physics in 1965 for his work on quantum electrodynamics (QED). In 1948, simultaneously with Schwinger and Tomonaga (with whom he shared the prize), he showed that QED can be renormalized, removing infinities, so that it gives finite results for all observable quantities. Schwinger and Tomonaga did this by building on existing theories in an ingenious way. Feynman invented a completely new method of treating the Schrödinger (or Dirac) equation. This was typical of Feynman's physics: he always had his own way of looking at and solving a problem.

Both methods were essential: Schwinger's immediately convinced quantum theorists, because it connected with previous knowledge. Feynman's seemed abstruse at first and Niels Bohr, for instance, found it hard to accept, but it soon became clear that it enormously simplified concepts and calculations. Only by these methods was it possible to extend theory to increasingly complicated problems and quantum electrodynamics to fabulous accuracy: the magnetic moment of the electron has now been calculated to an accuracy of 1 part in 10^8 , and measured to a similar accuracy.

Feynman also developed, in 1953, a fundamental theory of liquid helium, justifying the earlier theories of Landau and Tisza. Because 'He atoms are Bose particles, having spin zero, the ground-state wavefunction of a mass of liquid helium is symmetrical in all the particles and everywhere positive. There is only one such wavefunction, and the entire mass of helium behaves as one unit. This is why helium near 0 K is superfluid, showing no viscosity. At low temperature, pressure waves are the only possible motions in the liquid. At a higher temperature, about 0.5 K, it becomes possible for a small ring of atoms to circulate without other atoms being much disturbed; these are the 'rotons' of Landau theory. Feynman showed why the energy of a roton is minimum at a certain wavelength and that this wavelength is closely related to the average distance of helium atoms from each other.

To make a roton takes energy, therefore their number increases with temperature. They also interact with each other and thus show viscosity. An assembly of rotons therefore behaves much like a normal liquid, and moves independently of the superfluid. Helium may be regarded as a mixture of superfluid and normal liquid. When the concentration of normal liquid becomes too big, there is a phase transition in which the whole

liquid turns 'normal'.

Feynman's work on the nuclear weak interaction is also of fundamental importance. In 1956, Lee and Yang concluded that parity symmetry is not conserved in weak interactions, and this was soon confirmed experimentally by Wu and others. On the basis of experiments, physicists concluded that the violation of parity is the maximum possible. This stimulated Feynman and Gell-Mann to postulate that only the left-handed part of the wavefunction of a particle is involved in weak interactions. They postulated, in fact, that this was true not only for the neutrino, but for any particle, electron, muon and even composite particles like protons and neutrons. They also proposed that the weak interaction is universal: all particles interact and with the same strength. This theory permitted many conclusions, nearly all of them agreeing with experiments. One experiment at first seemed to disagree, but when repeated more carefully, agreed with the Feynman–Gell-Mann theory. The concept of chirality, left- or right-handed spin, turns out to be extremely fruitful in particle theory.

Experiments in the late 1960s on the scattering of high-energy electrons by protons at the Stanford Linear Accelerator showed very large inelastic scattering. Feynman soon concluded that smaller units were contained in the protons which he called partons, and that these collided elastically with electrons. The partons were soon identified with the quarks deduced from general theory, and one of the important tasks of particle physics has been the determination of the distribution

of quarks in proton and neutron. Feynman's attempts to understand quantum chromodynamics, the theory of parton interactions, were characteristically individual, especially his treatment of the confinement of quarks in hadrons. But his long struggle with abdominal cancer cut this work short.

Feynman enjoyed all life, and he lived in physics. He was able to transmit his enthusiasm to others, his colleagues and his students. He was beloved by my children, then 2–6 years old, for whom he babysat many times while he was at Cornell. It helped that he was also a clown. Once, when he was explaining nuclear fission at a high school, he did it simultaneously to two classes, standing in the doorway drawing pictures on two blackboards simultaneously with his right and left hands.

Feynman's approach was always direct, to life as to science. He disliked pompous people and made fun of them, but usually gently so as not to hurt them. His uncanny ability to get immediately to the core of a problem became well known when he sat on the commission to investigate the Challenger disaster and demonstrated the central problem simply by dropping a rubber O-ring into a glass of ice water.

In physics he always stayed close to experiment, and had no interest in esoteric theories. The three volumes, *The Feynman Lectures on Physics*, which tie the elementary problems of physics with the most advanced ideas, exemplify this. More than other scientists, he was loved by his colleagues and his students.

H.A. Bethe

H.A. Bethe is at the Newman Laboratory of Nuclear Studies, Cornell University, Ithaca, New York 14853, USA.

Molecular evolution

How old is a polymorphism?

Jonathan C. Howard

PERHAPS because the mathematics of classical population genetics entail the assumption of continuous interbreeding populations, polymorphism is often considered to be a within-species affair. If this were so, mutations which generate the allelic forms of polymorphic genes must have occurred within the lifetime of a single species, and variant forms of genes pre-existing in the ancestral population must not survive the speciation event. There is not a shred of evidence in favour of this position, and it is difficult to find a rational argument to support it with either. As all studies of wild populations show extensive heterozygosity at multiple loci, it is inconceivable that a founder population, no matter how small, will not transmit old alleles to new species. Speciation events

may well involve striking changes in allelic frequencies because of founder effects and new patterns of selective pressure, but that new species enter the world already polymorphic can hardly be disputed. This conclusion is not, of course, the same as saying that all polymorphism is ancient, only that some of it must be, and that arguments that depend on transmission of allelic forms through speciation events are not necessarily wrong.

On page 651 of this issue¹, McConnell *et al.* present direct evidence for trans-species polymorphism in the major histocompatibility complex (MHC) of the mouse. Even if the sound *a priori* argument above were not enough, several features of MHC polymorphism make trans-species transmission incomparably

more likely than not². First, many species of mammal have extraordinary numbers of apparently functional alleles segregating at several loci in the MHC, allelic frequencies tend to be low, and nearly all individuals are heterozygotes at most of these loci. Even very small founder populations would transmit several alleles at each segregating locus.

Second, the alleles of polymorphic MHC genes differ at multiple positions in the sequence. There is reason to believe that much of this sequence heterogeneity is initially driven into the population by strong selection, but in the absence of any evidence for a high mutation rate in this chromosomal region the allelic diversity of non-coding bases and intron sequences appears too high to be accommodated by neutral processes within the lifetime of a species.

Third, the role of MHC glycoproteins in the immune response suggests that individual alleles could have selective value that transcends the concerns of a single species. It is known³ that MHC glycoproteins collect processed antigenic proteins from intracellular compartments and present them as peptide fragments to the T-cell immune system. Conformationally labile peptide fragments of antigen of the order of 20 amino acids in length are presumably stabilized by binding into a deep cleft in the MHC glycoprotein lined by almost all the polymorphic residues⁴. Individual alleles probably represent more-or-less successful partial solutions to the general structural problem of being able to capture all peptides, and as such are of persistent value to any vertebrate immune system.

Structural marker

McConnell *et al.* have sought a direct test of the trans-species transmission of MHC polymorphism by identifying a structural marker in a non-coding region of one of the mouse MHC class II glycoprotein genes, A_{β} . The structural marker they found and describe in this issue¹ is a so-called SINE (short interspersed nucleotide element) sequence, a member of a family of repetitive elements which occur abundantly throughout the mouse genome. It is believed that SINE sequences are mobile within the genome, inserting randomly into chromosomal locations featuring suitable short recognition sequences. An intron that received a SINE sequence would thus be tagged for as long as that allele and its progeny remained in the population, and could be distinguished easily from alleles that did not contain the SINE sequence. The discovery of such a tagged allele segregating in two different species would strongly suggest that the original insertion event occurred in the common ancestor, and thus that the polymorphism must be older than either of the two present-day species.

McConnell *et al.* find a SINE sequence of 861 base pairs in the long intron between the β_1 and β_2 exons of the A_{β} gene. The SINE sequence clearly behaves like a polymorphic marker: some laboratory strains of the mouse *Mus musculus domesticus* possess the insertion (haplotypes *b*, *s*, *f*, *u* and *k*) whereas others lack it (haplotypes *d*, *q* and *nod*). McConnell *et al.* explore the genomic restriction maps surrounding the A_{β} gene in individual mice of several closely related species and suggest that the SINE sequence polymorphism is present in all these species as well. *Mus spretus* and *M. hortulanus* probably diverged from the lineage leading to the *M. musculus* species complex 3–4 million years ago. The natural conclusion is therefore that the A_{β} polymorphism tagged by the SINE insertion must be at least as old as this early branch point, and must have survived several speciation events.

Two further lines of evidence support the conclusions of McConnell *et al.*. If the A_{β} alleles of all mice, regardless of species, are grouped according to possession or lack of the SINE sequence, then other distinctive polymorphic features of the restriction maps outside the SINE insertion also fall into concordant groups. The implication is that SINE⁺ and SINE⁻ represent ancestral alleles which have evolved into two divergent clusters, each carrying markers that identify their common ancestry. In addition, comparison of the nucleotide sequences of the A_{β_1} and A_{β_2} domains of eight alleles of laboratory mice shows that some polymorphic sequence features are concordant with the grouping based on presence or absence of the SINE insertion.

Although the general picture presented by the results of McConnell *et al.* is consistent with the idea of an ancestral polymorphism tagged by an insertion more than 4 million years ago and preserved in the alleles of modern mice of various species, there are some peculiar aspects to the story. First, if MHC polymorphism was as extensive in the ancestral mouse species as it is in modern mice, and if transmission of polymorphism through speciation events is the rule rather than the exception, it is extraordinary that the SINE⁺ allelic group is apparently as homogeneous as the SINE⁺ group. The authors' argument depends on the SINE insertion having been a singular event which tagged a single allele. All other alleles would remain untagged. The SINE⁺ group should therefore represent the descendants of a single sequence, whereas the SINE⁻ group should represent an assemblage of all other alleles. It is, of course, possible that all SINE⁻ alleles except one got lost during the past 4 million years, but if so we have apparently returned to the position that most modern alleles must be of relatively recent ancestry. In fact, it is the SINE⁺ group which appears to have

internal complexity, based on polymorphism of the SINE sequence itself, restriction-site polymorphism, and polymorphic bases in the A_{β_1} and A_{β_2} domains. The authors argue that a second SINE insertion must have modified the first, and subsequently acted as a tag for the divergent evolution of a subset of the originally tagged group of alleles. Nevertheless, the apparent homogeneity of the SINE⁻ allelic group remains unresolved.

MHC allele evolution

Quite apart from the interest in establishing a measure for the age of a polymorphism, the work of McConnell *et al.* is also suggestive in relation to the pattern of MHC allele evolution. Spontaneous mutations at class I and class II loci in inbred strains of mice suggest⁵ that existing allelic sequence diversity may be increased by gene conversion involving short segmental exchanges between similar but non-identical sequences, generating new combinations of allelic markers. Although such sequence reshuffling must generate functionally distinct protein products, it does not, of course, introduce new mutations at any particular nucleotide. The evolutionary trajectory of any new single base change would, however, be hard to discern if sequence reshuffling between alleles occurred at a significant frequency. Indeed, one would anticipate that ancient allelic groups such as are described by McConnell *et al.* would not be visible. The data of these authors thus appear to suggest that reshuffling processes such as gene conversion are not important in sequence evolution of these genes in the wild.

On closer inspection, however, the sequence data collated and presented by the authors on this point is equivocal. Of the eight allelic A_{β_1} and A_{β_2} exon sequences shown, two (the *b* and *nod* alleles) have sequence features typical of more than one allelic group, implicating cross-over events in the generation of their characteristic sequences. Indeed, as shown in the figure, the whole of the authors' group 2, which is one of the two SINE⁺ groups, could be recombinant between the SINE⁻ group 1 (at the 5' end of the A_{β_1} exon), and the other SINE⁺ group 3 (the rest of the documented sequence). It could be significant that the integrity of the defined groups is much less convincing in the β_1 than in the β_2 exon for all the alleles. On the basis of the data presented, a good case could be made for sequence reshuffling affecting predominantly the β_1 exon. From the extensive human class I MHC allelic sequences collected⁶ by Parham *et al.*, a similar case could be made for gene conversion affecting the α_1 and α_2 exons, whereas α_3 seems to behave 'conventionally'.

Although the earliest divergence point studied by McConnell *et al.* is about 3–4

million years ago, the SINE insertion event could have been considerably earlier. The only rat A_{β} nucleotide sequence available⁷ should reveal whether the sequence motifs characteristic of the allelic groups of the genus *Mus* exist in a species whose time of divergence from the mouse must have been at least 10 million years ago (see figure). It is very striking how at almost every position (26/30) at which grouping according to McConnell *et al.* is defined, the rat gene shows one or other of the defining nucleotides. Admittedly, similar information is not available from another rat A_{β} allele, but it is tempting to propose that many of these substitutions are ancient, and will be found to have antedated rat-mouse divergence.

One of the biggest problems in long-range evolutionary analysis of MHC sequence variation follows from the instability of the genomic organization of both class I and class II genes, with extensive duplication and deletion. General family relationships are preserved, as between human HLA *DQ* and mouse *H-2 A*

will tend to enjoy a selective advantage. Frequency-dependent selection will therefore favour multiple alleles with low individual allelic frequencies, and hence a high degree of heterozygosity. The data presented by McConnell *et al.* suggest that at least a part of the polymorphic variation available in existing populations of mouse is relatively ancient, though this is probably more true of the substitutions themselves than of the array of substitutions that make up each modern allele, at least as far as the $A_{\beta 1}$ exon is concerned. The pattern of allelic evolution in the MHC emerging from this and other studies is consistent with the idea that several processes with different timescales are at work. There is no reason to invoke point hypermutation targeted to MHC genes.

However, new coding substitutions in regions of the molecule associated with peptide binding seem to be under intense selection. The low allelic frequencies suggest that while such new substitutions may enter the population rapidly, they rarely completely replace existing alleles.

Daedalus

Unspoken words

MACHINE-translation of speech into written text is still defeated, in all but the simplest cases, by the variety, subtlety and rapid variation of human speech-sounds. So Daedalus is devising a speech-writing system that ignores sound altogether. Instead, it follows the speech-mechanism that makes the sound. In effect, it is the ultimate in lip-reading.

Intrepid DREADCO volunteers are gargling with magnetic fluids and applying red-iron-oxide magnetic lipstick, so that the speech-movements of their coated lips, tongue and mouth can be followed by a web of external pick-up coils. Less invasively, electrodes taped to their faces are recording the changing distributed resistances and myoelectric potentials of their speech; even non-contact electrodes can register useful changes of distributed capacitance.

The output gathered by these methods is logically 'upstream' of the speaker's sound-output, and closely linked to his internal verbal code. So it should be much easier to decode back into words — just as it would be easier to reconstruct the music of a piano by following the key movements rather than listening to the sound. With luck, a few key facial parameters will suffice to define a speech-state uniquely, and DREADCO's Speakwrite[®] need only record these. The first commercial model will probably feature an electroded mask held close to the speaker's face; but Daedalus hopes ultimately to devise some form of radio-denture or radio-toothbrace to monitor the key parameters from inside the mouth, and transmit them to the interpreter/word-processor unit nearby.

Thus the irksome barrier between speech and writing will be broken down: intermediaries like secretaries and keyboards will become redundant. The author or executive will talk directly to his Speakwrite. The text coming up on the screen will doubtless contain mistakes, but these will easily be rectified by coded commands, the verbal equivalents of keyboard corrections. As a telephone-cum-telex machine, it will enable speech at one end to emerge as text at the other. And, being independent of sound, it will do its work in the most impossible sonic conditions. It will faithfully record the whispered or merely mouthed speech of the clandestine eavesdropper or Trappist novelist; it will accept the words of the tank-commander or rolling-mill manager against the most deafening clangour; it will flawlessly interpret the quacking falsetto of the deep-sea diver in his helium-oxygen atmosphere, or the thick croaking of the sufferer with a really heavy cold. Even the deaf and dumb will be able to converse by Speakwrite.

David Jones

$A_{\beta 1}$

$A_{\beta 2}$

Group 1	<u>A</u> . <u>GG</u> . <u>GAG</u> . <u>A</u> . <u>C</u> . <u>GC</u> . <u>CCGG</u> . <u>GATC</u> . <u>G</u> . <u>C</u> / <u>G</u>	<u>C</u> . <u>G</u> . <u>A</u> . <u>A</u> . <u>AT</u> . <u>A</u> . <u>C</u> . <u>T</u> . <u>A</u>
Group 2	<u>A</u> . <u>GG</u> . <u>GAG</u> . <u>G</u> . <u>T</u> . <u>AG</u> . <u>---</u> . <u>T</u> . <u>C</u> . <u>---</u> . <u>G</u> . <u>G</u>	<u>T</u> . <u>A</u> . <u>C</u> . <u>G</u> . <u>GG</u> . <u>G</u> . <u>C</u> . <u>T</u> . <u>A</u>
Group 3	<u>C</u> . <u>CC</u> . <u>TTC</u> . <u>A</u> . <u>T</u> . <u>AG</u> . <u>---</u> . <u>T</u> . <u>C</u> . <u>---</u> . <u>A</u> . <u>A</u>	<u>T</u> . <u>A</u> . <u>C</u> . <u>G</u> . <u>GG</u> . <u>G</u> . <u>T</u> . <u>C</u> . <u>C</u>
Rat	<u>A</u> . <u>GG</u> . <u>CTG</u> . <u>G</u> . <u>C</u> . <u>GC</u> . <u>AAGG</u> . <u>GATC</u> . <u>A</u> . <u>A</u>	<u>C</u> . <u>A</u> . <u>C</u> . <u>G</u> . <u>GG</u> . <u>G</u> . <u>C</u> . <u>A</u> . <u>C</u>

Diagnostic nucleotides defining the three groups of A_{β} alleles of McConnell *et al.* for the $A_{\beta 1}$ and $A_{\beta 2}$ domains, compared with the one available A_{β} sequence for the rat⁷. Only the group-diagnostic nucleotides are shown and other sequence is indicated by dots. Dashes indicate deleted nucleotides. Underlined nucleotides show identity with the rat sequence at these positions. The segmental organization of diversity is apparent in group 2, which begins with six bases from group 1, and thereafter follows group 3. Similarly, the rat sequence stays with group 1 for most of the $A_{\beta 1}$ domain, but follows group 3 thereafter.

genes, but there is no simple one-to-one correspondence between the human and mouse genes within the family. It is therefore impossible to follow allelic sequence variation with any assurance over such time intervals. In the rat, however, the central regions of the class II system seem to be orthologous to those of mouse, containing exactly the same genes in the same orientation in a simple one-to-one correspondence⁸. There is therefore no objection to extending the analysis¹ to the rat.

The selective forces responsible for MHC gene polymorphism are not fully understood, but probably follow from defensive strategies available to rapidly evolving pathogens. To avoid T-cell recognition and immunity, potentially antigenic peptides of pathogen origin either must fail to bind to MHC molecules, or in doing so must mimic self structures to which the host is tolerant. In either case, possession of two different allelic forms of MHC molecule improves the host's chances of mounting a successful immune response. Conventional heterozygous advantage is therefore probably involved. Second, pathogen evolution will be guided by the need to subvert abundant allelic forms, so that rare alleles

Rather, they join the very large pool of alleles already present and stay fluctuating in frequency in the population for several million years. During this time, allelic point mutations are reshuffled with others by recombinational mechanisms to generate functionally new alleles which are themselves subject to further rounds of pathogen-dependent selection. Perhaps peptide-binding domains of MHC molecules are favoured targets for reshuffling processes, and other parts of the genes evolve fairly conventionally. Only by further fine analysis of the structure of MHC sequence polymorphism within and between species will the relative importance of these distinct mechanisms become clear. □

1. McConnell, T. J., Talbot, W. S., McIndoe, R. A. & Wakefield, E. K. *Nature* **332**, 651–654 (1988).
2. Klein, J. *Hum. Immun.* **19**, 155–162 (1987).
3. Germain, R. N. *Nature* **322**, 687–689 (1986).
4. Bjorkman, P. J. *et al.* *Nature* **329**, 506–518 (1987).
5. Nathanson, S. G. *et al.* *A. Rev. Immun.* **4**, 471–502 (1986).
6. Parham, P. *et al.* *Proc. natn. Acad. Sci. U.S.A.* (in the press).
7. Eccles, E. J. & McMaster, W. R. *Immunogenetics* **22**, 653–663 (1985).
8. Diamond, A. G. *et al.* *Transplant. Proc.* **17**, 1808–1811 (1985).

Jonathan C. Howard is in the Department of Immunology at the AFRC Institute of Animal Physiology and Genetics, Babraham, Cambridge CB2 4AT, UK.

***Lycopodium* spores found in condom dusting agent**

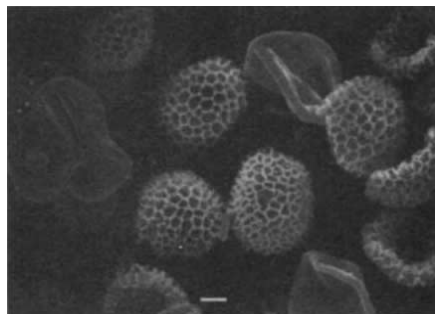
SIR—Following an inquiry from an industrial supplier of botanical products, the use of a potentially hazardous plant product in the production of condoms has come to our attention. The product, spores of the staghorn clubmoss *Lycopodium clavatum* L. (Lycopodiaceae) — also called vegetable sulphur¹ — is used as a dusting agent on at least one major brand of non-lubricated condoms to prevent the rolled latex from sticking to itself.

While trying to locate additional supplies of this product, we discovered numerous references to its use and hazardous nature. The spores contain 50 % of a fixed oil, as well as sucrose and phyto-sterol^{2,3}. Due to the high proportion of oil, the spores are resistant to wetting, have emollient properties, and are highly flammable. Spores of *Lycopodium* species, including *L. clavatum*, have been used for a variety of purposes. Exposure to *Lycopodium* spores can cause allergic reactions ranging from dermatitis to severe asthma attacks. These reactions have been recorded in pharmacists exposed during preparation of spore-coated pills and suppositories; in metal workers exposed to dry parting compound used to prevent metal from sticking to wooden moulds; in theatre personnel and patrons because of exposure to face and hair powders; as a hay-fever response to natural exposure; and in consumers using dry shampoos, face and hair powders, and spore-coated pills^{3,4}.

Of most concern are reports of *Lycopodium* spores causing adhesions on serous surfaces and foreign-body granulomas in soft tissue through introduction into wounds from rubber gloves and supplies used during surgery^{3,5-8}. A foreign-body response causes lesions composed of granulation tissue arranged in the form of nodules, consisting of epithelioid cells, multinucleated giant cells, lymphocytes and fibroblasts with areas of necrosis. It simulates neoplastic disease or infective granulomas caused by bacteria such as tuberculosis or syphilis. The causative agent may be overlooked because the spores are poorly visualized in haematoxylin and eosin, but stain bright red and acid-fast in Ziehl-Neelsen preparation. In animal experiments, granulomas are formed within 2–6 weeks after exposure to *Lycopodium* spores⁶. One of the more disturbing case histories was a report of *Lycopodium* granuloma caused by the use of spore-covered anal suppositories that came in contact with a wound from an external haemorrhoidectomy⁷.

By scanning electron microscopy, spores ("Lycopodium, Canadian") provided by the industrial supplier as typical of those used as a dusting agent on condoms, appeared identical to spores of *L.*

clavatum, although several species of *Lycopodium* found in Canada have similar spore ornamentation^{9,10}. Surveying three commonly sold brands of condoms at a local pharmacy, one (Ramses, non-lubricated) was identified as having been treated with *Lycopodium* spores. Microscopy of the condom surfaces reveals numerous irregular polyhedrons of varying sizes. Some of the polyhedrons, which disappeared after acetolysis, correspond to the typical morphology for grains of corn starch. As shown in the figure, some of the spores which remain after acetolysis appear identical to *L. clavatum* spores, while others are an unidentified larger spore with smooth walls. Because the spores are on the inside and outside surfaces of the condom, they would come in contact with membranes, both mucous and non-mucous, during use.



Scanning electron micrograph (scale bar, 10 µm) of dust-like material from condom, after acetolysis; view of *Lycopodium* spores (reticulations) and unidentified pteridophyte spores (smooth walls). Micrograph by Donald Black.

As condom use is increasing dramatically in response to the AIDS epidemic, physicians should take note of the possibility that granulomatous masses on any of the areas of the body that come in contact with condoms might be traced to the use of these products. These granulomas are non-lethal, do not lead to cancer and are easily remedied; this is a relatively minor health problem compared with AIDS, a lethal sexually acquired disease that might be contracted by discontinuing the use of such protective devices.

MICHAEL J. BALICK*

JOSEPH M. BEITEL†

Institute of Economic Botany,
and Horticulture Department and
Herbarium†,
New York Botanical Garden,
Bronx, New York 10458, USA*

1. Windholz, M. ed. *Merck Index* (Merck and Co., Rahway, New Jersey, 1983).
2. Greenburg, L.A. & Lester, D. *Handbook of Cosmetic Materials* (Interscience, New York, 1954).
3. Whitebread, C. *Am. Fern J.* **31**(3), 100–102 (1941).
4. Salen, E.B. *Acta allerg.* **4**, 308–319 (1951).
5. Erb, I.H. *Surgery Gynec. Obstet.* **60**, 40–44 (1935).
6. Antopol, W. *Archs path.* **16**, 326–331 (1933).
7. Antopol, W. & Robbins, C. *J. Am. med. Assoc.* **109**, 1192 (1937).
8. Boyd, W. *Pathology for the Surgeon* Ch. 4 (Saunders, Philadelphia, 1955).
9. Wilce, J.H. *Am. Fern J.* **62**(3), 65–79 (1972).
10. Beitel, J.B. *Bull. Am. Fern Soc.* **6**(5), 1–8 (1979).

Myosin heads and assembly of muscle thick filaments

SIR—Small in his recent News and Views article¹ analysed new experimental findings concerned with myosin filaments that provide some explanations of the role of magnesium-ATP in filament assembly. But as long ago as 1979, it was found that MgATP is involved in the regulation of the diameter of skeletal muscle myosin filaments². The authors found that, in the absence of MgATP, synthetic filaments are 30–50 nm wide compared with 16 nm for native filaments. In the presence of millimolar concentrations of MgATP, the diameter is 16 nm, indicating the crucial role of this nucleotide in filament assembly. It was also shown that free ATP disrupts the filaments². These authors, however, did not realize the significance of MgATP. But in 1985, Pinset-Härström³ wrote: "the MgATP interaction sites involved are the high-affinity sites located in the heads, and [this] is to my knowledge the first indication that these sites are implicated in skeletal filament assembly". Other work, however, indicated⁴ that the high-affinity sites are not involved in filament structure.

According to Small¹, "even more intriguing are the effects of myosin heads on filament assembly and disassembly". In 1982, we published our hypothesis, according to which one head of a myosin molecule lies outside the filament core while the other is buried in the backbone and interacts with the internal head of the opposite molecule. This hypothesis⁵ was based on our discovery of a head dimer⁶ whose properties depend on the same parameters as the filaments: strong interaction of the subunits in the presence of MgATP and MgADP; and disappearance of the dimer in the presence of free ATP, free ADP or at high ionic strength.

The head dimer has been seen directly in the electron microscope, both in S1 and on intact myosin filaments. Its existence, therefore, must be accepted. We believe that the problem of the myosin heads is about to be elucidated. The involvement of the heads in filament assembly is becoming generally accepted, and it is now important to discover their exact role.

JEAN E. MOREL

NACIMA BACHOUCHI

*Département de Biologie,
Service de Biophysique,
CEN-Saclay, 91191 Gif-sur-Yvette Cédex,
France*

1. Small, J.V. *Nature* **331** 568–569 (1988).
2. Pinset-Härström, I. & Truffey, J. *J. molec. Biol.* **134**, 173–188 (1979).
3. Pinset-Härström, I. *J. molec. Biol.* **182**, 159–172 (1985).
4. Chowrashi, P.K. & Pepe, F.A. *J. Muscle Res. Cell Motil.* **7**, 413–420 (1986).
5. Morel, J.E. & Garrigos, M. *FEBS Lett.* **149**, 8–16 (1982).
6. Morel, J.E. & Garrigos, M. *Biochemistry* **21**, 2679–2686 (1982).

Cometary organic matter still a contentious issue

SIR—Hoyle and Wickramasinghe¹ question certain aspects of our model of the 3.4- μm feature in the coma of comet Halley as thermal emission from organic grains². Our basic contention is that if the infrared spectra of freeze-dried bacteria do acceptably match the cometary spectra—a matter in dispute—this is because of simple organic functional groups that are present in many abiogenic organics. We therefore do not believe, as Hoyle and Wickramasinghe charge, that our preference for a non-biological explanation is due to our “strongly-held cultural prejudice that there is no life outside the Earth”. (Indeed, one of us has made a considerable effort over three decades to seek evidence relevant to such life.) We do concede, however, that we should not have described the fit between the bacterial model and the 3.4- μm feature as providing “no” evidence for cometary biology; rather, we should have said that it provides no persuasive evidence as much more plausible alternatives exist. The recent comparison³ of ‘bacterial’ and ‘abiotic’ models leaves unanswered such questions as how the bacterial emission model somehow fits the 3.05- μm absorption feature, or whether this model implies emission features at longer wavelengths where none is observed.

Hoyle and Wickramasinghe ask what “definite observation” we can cite for our “belief in the production of a specific mixture of organic materials in significant quantity” from irradiated C- or N-bearing ices. The production of organics from such ices has an experimental history extending over a quarter of a century^{4–6}, and was recently the subject of an international workshop⁷. The application of such experiments to the outer Solar System has recently been surveyed⁸.

Most points raised by Greenberg and Zhao⁹ with reference to our original correspondence¹⁰ are addressed directly in our subsequent letter²; indeed, several of their points are possibilities which we explicitly endorsed. They object to “the assumption that a methane ice clathrate has any relevance to the comet nucleus.” But recent work based on *in situ* data from the Giotto ion mass spectrometer indicates that CH_4 is present in the Halley coma with a production rate about 2% that of H_2O ¹¹. In cometary ices, clathrates are thermodynamically favoured¹². By no means, however, did we argue that CH_4 clathrate is the only plausible candidate ice yielding radiation-processed organics, and our group has examined a range of alternatives^{8,13}. However, interstellar dust particles of probable cometary origin provide “nothing resembling the Greenberg model”¹⁴ (see ref. 15). Nor did we claim that only post-accretion cosmic-ray

processing is important for the formation of organics. Rather, we concluded² that most organic Halley dust “is probably jetted from the interior, with organics thus due mainly to processing by radionuclides and pre-accretion irradiation”.

Greenberg and Zhao note that “no published laboratory residue spectrum provides a perfect fit to the Halley dust emission”. But no single model can yet hope to encompass the entire range of complexity of the formation and time-variable infrared emission of cometary organics. It is not even certain that the 3.4- μm feature is wholly due to thermal emission from organic grains², rather than, for example, gas-phase fluorescence. As is often the case in science, the present debate is not about which mechanism is ‘the’ answer, but rather about the relative merit of competing models.

Our model makes specific testable predictions for the heliocentric evolution of cometary emission features, to be tested by observations planned for comet Brorsen–Metcalf in 1989 (C.C. and C.S., preprint). We hope the infrared astronomy community will take full advantage of this apparition. Only further experimental work, tested against such observations, can sift through the range of alternative models.

CHRISTOPHER CHYBA

CARL SAGAN

Laboratory for Planetary Studies,
Cornell University,
Ithaca, New York 14853, USA

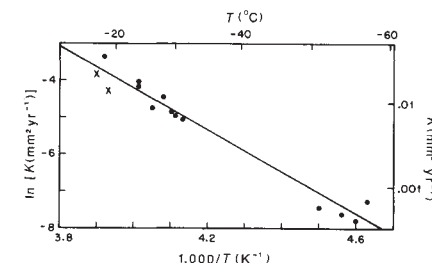
1. Hoyle, F. & Wickramasinghe, N.C. *Nature* **331**, 123–124 (1988).
2. Chyba, C. & Sagan, C. *Nature* **330**, 350–353 (1987).
3. Hoyle, F. & Wickramasinghe, N.C. *Nature* **331**, 666 (1988).
4. Berger, R. *Proc. natn. Acad. Sci. U.S.A.* **47**, 1434–1436 (1961).
5. Khare, B.N. & Sagan, C. in *Molecules in the Galactic Environment* (eds Gordon, M. & Snyder, L.) 399–408 (Wiley, New York).
6. Greenberg, J.M. *et al. J. phys. Chem.* **87**, 4243–4260 (1983).
7. *Proc. Int. Workshop on Laboratory Simulation of Organic Cometary Material*, eds Kissel, J. & Strazzulla, G. (University of Catania, 1987).
8. Thompson, W.R. *et al. J. geophys. Res.* **92**, 14, 933–14,947 (1987).
9. Greenberg, J.M. & Zhao, N. *Nature* **331**, 124 (1988).
10. Chyba, C. & Sagan, C. *Nature* **329**, 208 (1987).
11. Allen, M. *et al. Astron. Astrophys.* **187**, 502–512 (1987).
12. Delsemme, A.H. *Mem. Soc. R. Sci. Liège* **9**, 135–145 (1976).
13. Khare *et al. Icarus* (submitted).
14. Walker, R.M., in *Infrared Observations of Comets Halley and Wilson*, NASA Conf. Publ. (ed. Hanner, M.) (in the press).
15. Brownlee, D.E., in *Infrared Observations of Comets Halley and Wilson*, NASA Conf. Publ. (ed. Hanner, M.) (in the press).

Long-term climate changes from crystal growth

SIR—In a recent letter to *Nature*¹, Petit *et al.* proposed that the small grain sizes in Wisconsinan ice from Vostok Station and Dome C, East Antarctica, were caused by the cold surface temperatures at the time of deposition. A further analysis of the relevant data indicates that their proposed temperature ‘memory’ can last only hours rather than thousands of years, and that

the small grain sizes in Wisconsinan ice are probably caused by the drag effect of the soluble impurities in that ice².

Petit *et al.*¹ proposed that the grain-growth rate is controlled by the concentration of self-interstitials in the ice lattice near grain boundaries, and that this concentration becomes fixed soon after deposition and does not change until the onset of recrystallization near the glacier bed. Formation of self-interstitials is thermally activated, and free surfaces (bubbles or pores) serve as sources or sinks for self-interstitials during temperature changes³. The proposal of Petit *et al.*¹



Temperature dependence of grain growth in polar ice. Impure coastal sites are shown by crosses. Data include all points from ref. 1, plus further points from refs 5 and 9.

thus requires that the diffusion distance for self-interstitials over time $t \approx 10^4$ yr is small compared with the spacing between pores ($\approx 10^{-3}$ m), so that equilibrium is not established. At -60°C (the approximate temperature of Vostok and Dome C) the diffusion coefficient for self-interstitials is $D = 3 \times 10^{-10} \text{ m}^2 \text{ s}^{-1}$ (ref. 3). Estimating the diffusion distance as $(Dt)^{1/2}$ excess interstitials will diffuse 10^{-3} m in about 1 h. (A more exact calculation for diffusion to a spherical bubble⁴, using data for Wisconsinan ice from Dome C, shows that excess interstitials quenched into ice by an instantaneous cooling from -10°C to -60°C would be reduced to less than 1 per cent of the equilibrium concentration at -60°C in less than a day.)

A similarity between the activation energy for volume diffusion in ice and that for grain growth, calculated from an Arrhenius plot of growth rates at different polar sites, is cited as evidence¹ that volume diffusion of self-interstitials controls grain growth. But similarity of activation energies is not sufficient to demonstrate equality of mechanisms. Furthermore, reanalysis of available data shows that the activation energies are significantly different in this case. Petit *et al.*¹ combined data corrected for the section effect (the difference between the average cross-sectional area of a grain on a plane of section and the true cross-sectional area) with uncorrected data. The figure shows an Arrhenius plot similar to that of Petit *et al.*¹, in which available data have been corrected in a geometrically consistent manner to match Gow⁵ as closely as

possible. This gives an activation energy of 0.49 ± 0.03 eV, significantly lower than estimates of the activation energy for volume diffusion of 0.60 eV (ref. 3) and 0.62 ± 0.04 eV (ref. 6).

Petit *et al.*¹ discount the possibility that dissolved impurities slow grain growth² by noting that coastal sites with high concentrations of soluble impurities (especially NaCl) plot near (although slightly below) the same Arrhenius trend as inland sites with lower concentrations of impurities. But the coastal sites^{1,4} at -19 and -17 °C are warmer than the NaCl–H₂O eutectic (-21 °C). With increasing temperature above the eutectic, impurity drag is reduced to zero and enhanced boundary migration begins during a transition from solid- to liquid-phase boundaries. Experimental results from Jellinek and Gouda⁷ show that this transition occurs between -23 and -10 °C at 10^{-2} M NaCl, and results from de Achaval *et al.*⁸ show that 10^{-2} – 10^{-4} M NaCl speeds grain growth at -16 °C and above; natural concentrations of soluble impurities are 10^{-5} – 10^{-6} M in coastal sites and 10^{-6} – 10^{-7} M inland¹. Therefore, it is possible that inland sites experience significant impurity drag, but that warm coastal sites are near the transition and experience little impurity drag; further study is warranted. The temperature 'memory' proposed by Petit *et al.*¹ thus cannot explain the small grain sizes in Wisconsin ice, but available data are consistent with an impurity-drag model.

R.B. ALLEY*
J.H. PEREPEZKO†
C.R. BENTLEY*

*Geophysical and Polar Research Center
and

†Materials Science Program,
University of Wisconsin,
Madison,
Wisconsin 53706, USA

PETIT *ET AL.* REPLY—In polar deep ice cores, the climatic change between the Holocene and the last glacial maximum, as depicted from the stable isotope content, is accompanied by modifications in crystal size and impurity content. We have interpreted¹ the change in crystal growth rate mainly as a result of a climatic (temperature) effect whereas Alley *et al.* (ref. 2 and above) attribute it to the effect of soluble impurities such as NaCl. Here we discuss the stability of extrinsic interstitial defects, the value of the activation energy for the growth process, and the effect of the soluble impurity content.

First, our model basically assumes a built-in memory of the thermal conditions prevailing in the top layers of the firn through defect formation in the ice crystal lattice. Our assumption concerning the formation of extrinsic defects through an ageing effect in the top few metres of snow is supported by several physical measurements^{10,11} (dielectrical and micromechanical

damping behaviour of ice). These extrinsic defects (interstitials or more complex defects) are likely linked to structural defects such as sub or grain boundaries¹¹. The mobility of these defects could control the migration of boundaries¹. Alley *et al.* discuss the diffusion coefficient that concerns the diffusion of self-interstitials, but not the diffusion of interstitials associated with boundary migration. The structure of ice crystals after ageing could correspond to a lower free energy¹¹. It should therefore be stable as long as there is no nucleation of new crystals during dynamic recrystallization, and is preserved in deep-ice core records.

Second, Alley *et al.* obtain 0.49 eV for the activation energy deduced from the Arrhenius plot of available data. This value differs from ours because we excluded the data from Plateau Station (-57 °C) because of its highly uncertain chronology. In any case, with a value of 0.49 eV, the temperature change between the Holocene and the last glacial maximum estimated by our model would be only slightly affected, increasing by about 15 per cent.

Third, Alley *et al.* assert that above the NaCl–H₂O eutectic temperature (about -21 °C), the impurity drag effect is reduced to zero with the formation of a liquid phase at grain boundaries. But if we assume, as did Alley *et al.*², that Na⁺ and Cl[−] dissolve in the ice lattice and that grain boundaries are not saturated, the formation of the eutectic below -21 °C and of a liquid phase above would not be expected. Recent work¹² on the localization of impurities in a polar ice sample with 320

parts per 10⁹ (p.p.b.) Cl[−] content (Dollman, Antarctica) shows that no Cl (less than 1 per cent) is at triple junctions, even though the temperature *in situ* is as high as -16.7 °C. Any liquid present in polycrystalline ice should be preferentially at triple junctions¹³. As grain growth data from coastal Holocene ice at -19 °C and -17 °C (with 200–400 p.p.b. Na) are in accordance with data from inland sites at lower temperatures (with 20–40 p.p.b. Na), the drag impurity effect cannot be used to explain grain growth in polar ice¹.

In conclusion, on the basis of the limited data available, our interpretation remains valid, but the grain growth observations from sites with very different impurity concentrations do not support the drag impurity model of Alley *et al.*².

J.R. PETIT
P. DUVAL
C. LORIS

Laboratoire de Glaciologie et
Géophysique de l'Environnement,
B.P. 96, 38402 St Martin d'Heres Cedex,
France

1. Petit, J.R., Duval, P. & Loris, C. *Nature* **326**, 62–64 (1987).
2. Alley, R.B., Perepezko, J.H. & Bentley, C.R. *J. Glaciol.* **32**, 415–424 & 425–433, (1986).
3. Goto, K., Hondoh, T. & Higashi, A. *Jap. J. appl. Phys.* **25**, 351–357 (1986).
4. Ham, F.S. *J. Phys. Chem. Solids* **6**, 335–351 (1958).
5. Gow, A.J. *IAHS Publ.* **114**, 25–40 (1975).
6. Hobbs, P.V. *Ice Physics* (Clarendon, London, 1974).
7. Jellinek, H.H.G. & Gouda, V.K. *Phys. Status Solidi* **31**, 413–423 (1969).
8. de Achaval, E.M., Nasello, O.B. & Ceppi, E.A. *J. Phys., Paris* **48**(C1), 283–288 (1987).
9. Alley, R.B. thesis, Univ. Wisconsin (1987).
10. Vassaille, R. *et al. Ann. Geophys.* **36**, 491–498 (1980).
11. Tatibouet, J. *et al. J. Physique* **42**, 1473 (1981).
12. Mulvaney, R. *et al. Nature* **331**, 247 (1988).
13. Nye, J.F. & Frank, F.C. *Proc. Symp. IAHS Publ.* **95**, 157–161 (1973).

Sexual selection and predation risk in guppies

SIR—Breden and Stoner¹ present the results of an experiment apparently showing that predation risk determines female preference of male colour patterns in wild guppies; guppies from places with low predation apparently prefer brighter model fish whereas those from high predation apparently prefer duller models. Apart from the general implications for sexual selection theory, these results could explain the remarkably rapid increase to predation in my own experiments², because sexual selection and predation would both favour less conspicuous colour patterns under high predation. Unfortunately, there are some serious flaws in the experimental design, as well as some problems of interpretation, which make the interpretation concerning sexual selection questionable. Here, I point out the difficulties of experiments of this nature and suggest ways in which to demonstrate this potentially very interesting phenomena.

In Breden and Stoner's experiment¹, the males to be chosen by females were models made from painted balsa wood,

and were 4.8 cm total length by 1.3 cm high, which is as much as four times larger than real males^{3,4}. Therefore, female guppies may respond to these models not as super-normal guppy stimuli but as potential predators such as *Rivulus*, *Hemibrycon* and *Astyanax*, which are about the same size^{3,5} as the models.

In addition, Breden and Stoner placed an aerator near the brighter model, and not near the dull one so the brighter one moved more. It is unfortunate that this experiment was not properly controlled with both models moving equally; females from high-predation stocks could have been more frightened of the brighter, more rapidly moving models, and so 'preferred' the duller models. Females from the low-predation stocks were less frightened by the apparent predators (as they normally do not experience predation^{3,5,6}), and may even have been attracted by the colours. This is consistent with several studies of fright and other anti-predator behaviour in guppies, which is also heritable (refs 7–10, and P.H. Luyten, unpublished thesis). Therefore, the results

of Breden and Stoner are consistent with genetic variation for fright behaviour in the presence of potential predators and may have nothing to do with sexual selection.

The results¹ are also open to different interpretations not involving sexual selection. Guppies are known to have a preference for colour patterns of their own population, and to discriminate against those of others, even within the same stream. Given a choice between a native and an alien colour pattern, a female usually chooses the native colour pattern^{11,12}. Breden and Stoner's design was a comparison of colour patterns characteristic of different streams (high and low predation). Assuming that the models are not interpreted as potential predators, females from high-predation stocks would prefer the duller models as they are more similar to colour patterns in their own population, and the brighter models are not recognizable as males. At the same time, the females from the low-predation stocks would prefer the brighter model because they are more similar to their own patterns. This has nothing to do with sexual selection evolving in response to predation. To demonstrate a significant preference for duller males in females from high-predation stocks, it is necessary to give such females choices between males near the extremes of the natural range of variation within the same population, thereby eliminating discrimination against foreign colour patterns. The same must be done within the low-predation stocks, as it has in other experiments^{6,13}.

Guppies vary genetically with predation intensity in two aspects of behaviour, schooling^{3,7-10} and courtship mode^{9,10}. The naturally occurring sex ratio is female-biased³⁻⁶, so most fish in a school are not coloured. Females are attracted to other fish as part of the schooling response, and fish from high-predation areas have a greater tendency to be attracted to other fish than low-predation stocks. The less-coloured model may therefore seem more like a fish to school with than the bright model, and the bright model may seem more like a predator. This also yields the observed results¹ with no reference to sexual selection. In addition, guppies from low-predation areas court more vigorously and indulge in more visual displays than guppies from high-predation areas, and these differences are heritable^{9,10}. As a result, females from low-predation areas may expect or require more male movement before responding to models than females from high-predation localities. Again, this yields the observed results¹ without sexual selection.

Furthermore, degree of brightness is specifically dependent upon the background and visual conditions^{6,14} which vary radically between natural streams and populations. These parameters were not

controlled in Breden and Stoner's experiment, yet they are known to influence female behaviour⁶. 'Brightness' is not a simple one-dimensional property, and must be measured directly in a way which is meaningful to the signal receiver before any conclusions can be made about sexual selection for brightness.

J.A. ENDLER

*Department of Biological Sciences,
University of California,
Santa Barbara, California 93106, USA*

BREDEN REPLIES—Our study¹ correlated genetic differentiation for female preference in the Trinidad guppy with variation in the strength of selection against the attractive male character, as predicted by fisherian models of sexually selected female preference¹⁵. Endler has three alternative explanations for our results: (1) that females were attracted to familiar male patterns; (2) that females were responding to the models as if they were predators; and (3) that female mating preference measured independently of male mating success might reflect preference related to other behaviours. None of these alternatives is correct, and the question of familiarity may reflect a misunderstanding of the relationship between our results and the theory of the evolution of female preference by sexual selection.

If Endler is referring to a familiarity caused by females interacting with males during development, then I point out that, as stated in our paper, we tested naive females. Offspring were never exposed to their male parent, and all male offspring were removed before they displayed either secondary sexual coloration or courtship behaviour. If Endler is using the term familiarity to mean a genetic predisposition for patterns present in the native population, then that is exactly what we examined. We suggest that an attractive male character and female preference for it are coevolving.

At this stage in our studies, we have concentrated on total brightness, using coloration as an attractive character known to respond to natural selection on males^{2,5}. Because of this question of total brightness versus specific male patterns, we constructed models that were not present in the natural populations. The bright model incorporated spots and stripes characteristic of many different bright males from many populations, but it did not comprise any specific male pattern or any pattern typical of a natural population. The drab model was uniformly dull yellow, although drab males in nature show some coloration. By placing the aerator closer to the bright model, we correlated movement and brightness, because male movement and brightness contribute both to increased male mating success^{13,16} and to increased risk of predation^{2,5,16}. Further experiments will allow us

to examine what features of the attractive model are necessary to elicit female response, similar to previous studies that used synthetic, attractive signals to determine female preference¹⁷. Thus, we can determine if components of the attractive signal correspond to components of naturally occurring male patterns.

The females in our study did not respond as if they perceived the models to be predators. First, as we stated, females responded in a fashion similar to those tested using live males as targets¹⁸, including performance of the 'move-toward' behaviour assumed to be part of female sexual response¹³. On average, 6.21 of the 20 observation intervals for high-predation females tested with live males¹⁸ included a visit to the column of water nearest the bright male, while 5.02 of these intervals for high-predation females tested with models¹ included a visit to the bright model ($P > 0.05$). It seems inconceivable that females would approach as closely as they did the bright model or would respond to it in a manner so similar to responses to real males if they perceived the model to be a predator. Second, the fact that some females from high-predation populations actually showed a preference for the bright model argues against the idea that they perceived it to be a predator.

In general, studies of female preference that rely on measuring choice independently of measuring male mating success cannot exclude alternative preferences (for example, for schooling with a certain type of individual). However, in the guppy system male coloration affects mating success^{2,5}, and female preferences measured by male mating success has been correlated with preference measured by female behaviour alone¹³. This correlation, together with the potential to manipulate one variable of the synthetic, attractive signal while controlling all others, means that the use of models is a powerful method for determining the full range of female-preference functions.

F. BREDEN

*Division of Biological Sciences,
University of Missouri,
Columbia, Missouri 65211, USA*

1. Breden, F. & Stoner, G. *Nature* **329**, 831–833 (1987).
2. Endler, J.A. *Evolution* **34**, 76–91 (1980).
3. Liley, N.R. & Seghers, B.H. in *Function and Evolution in Behaviour* 92–118 (Clarendon, Oxford, 1975).
4. Reznick, D. & Endler, J. *Evolution* **36**, 160–177 (1982).
5. Endler, J.A. *Evol. Biol.* **11**, 319–364 (1978).
6. Endler, J.A. *Envir. Biol. Fish.* **9**, 173–190 (1983).
7. Seghers, B.H. *Evolution* **28**, 486–489 (1974).
8. Breden, F. *et al. Anim. Behav.* **35**, 618–620 (1987).
9. Crow, R.T. thesis, Univ. British Columbia (1981).
10. Luyten, P.H. & Liley, N.R. *Behaviour* **95**, 164–179 (1985).
11. Crow, R.T. & Liley, N.R. *Can. J. Zool.* **57**, 184 (1979).
12. Haskins, C.P. & Haskins, E.F. *Proc. natn. Acad. Sci. U.S.A.* **36**, 464–476 (1950).
13. Houde, A.E. *Evolution* **41**, 1–10 (1987).
14. Endler, J.A. in *Predator-Prey Relationships* (eds. Feder, M.E. & Lander, G.V.) 109–134 (Univ. Chicago Press).
15. Lande, R. *Evolution* **36**, 213–223 (1982).
16. Bischoff, R.J. *et al. Behav. Ecol. Sociobiol.* **17**, 253 (1985).
17. Klump, G.M. & Gerhardt, H.C. *Nature* **326**, 286 (1987).
18. Stoner, G. & Breden, F. *Behav. Ecol. Sociobiol.* (in the press).

Deep into the shallows

J.H.S. Blaxter and A. Edwards

Georges Bank. Editor-in-chief Richard H. Backus. MIT Press: 1987. Pp. 593. \$225, £160.95.

GEORGES Bank is situated off the coast of New England, with its western edge about 120 km from Cape Cod. It measures about 280 km by 150 km, of which some 14,000 km² is less than 60 m deep. At its northern and western edges it is connected by sills to the coasts of Canada and the United States, and so encloses the deeper waters of the Gulf of Maine.

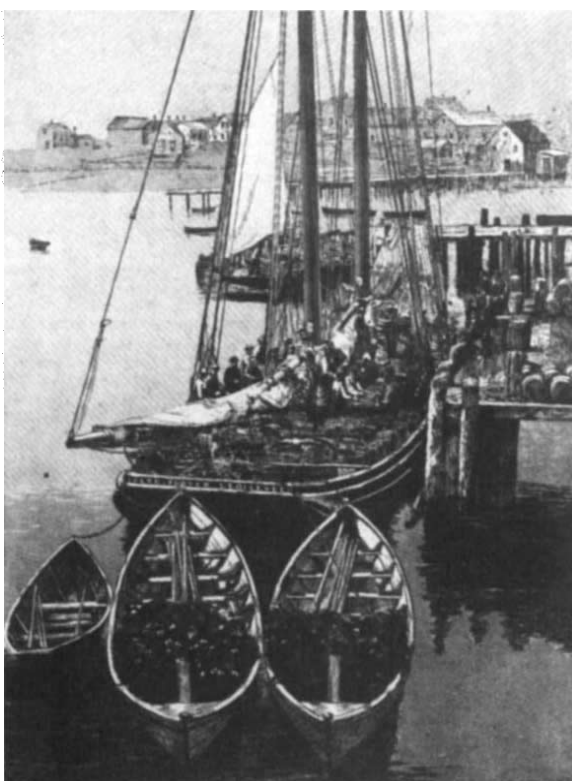
Georges Bank has been an important fishing ground since the middle of the eighteenth century. Even in the nineteenth century signs of overfishing were evident, and unexplained ups and downs in the yield have meant varied fortunes for fishermen from both the New and the Old Worlds. In modern times, about 20 species from a total of about 100 have contributed to annual yields as high as 500,000 tonnes. Since the 1970s, however, a steep and consistent reduction in the catch followed heavy fishing, fleets from the USSR, Poland, and East and West Germany all taking a major share. These foreign fleets have now departed, leaving a depleted stock to the United States and Canada.

Although the Bank has always mattered to New Englanders because of its fisheries, much wider interest was sparked off by US plans in 1974 to lease parts of it for oil exploration. Lobbying by vested interests, especially the fishing industry, delayed the first drilling until July 1981. Exploitation was also hindered by the extension by both Canada and the United States of their fishing limits to 200 miles in 1977. The management zones overlapped, the conflict between the two countries was never amicably resolved, and a settlement was finally reached only after a hearing by the International Court of Justice in 1984.

This is the background to the book under review — a massive, multi-author work of 57 articles and 593 pages, each measuring 39 cm × 34 cm. Even so, it is an effective condensation of reality, using only 10 cm² of paper per square kilometre of bank. A bargain billion fold reduction!

In 1979, when the dispute over Georges was at its hottest, the Woods Hole Oceanographic Institution set up a coastal research centre. One of its remits was a long-term study with a view to finding out

why the Bank was so productive. This book was conceived as a project summary, to treat all aspects of the natural science of the Bank, to describe its resources and to consider some of the political issues associated with their exploitation. Although



Been fishing — a purse seiner with barrels of salted mackerel on deck, a picture of 1887.

technical in content, it was intended to inform not only scientists but also students, legislators, fishermen, engineers, lawyers and resource managers.

A first impression is that the editors have achieved their aim, a judgement that stands up to closer scrutiny. The feast of articles consists mainly of scientific reviews (although we are told that new material is also included) together with good accounts of history, commerce, economics, politics, and there are even some fishing anecdotes to leaven the fairly substantial meal. The style is in places somewhat more technical than *Scientific American* but the general appearance of the volume, with photographs, figures and maps, reminds one of that publication. The quality of production is extremely high.

The book's gestation period of seven years comes as no surprise to anyone asso-

ciated with the publication of multi-author works. The contributors involved themselves in a range of informal workshops, 'tutorial' lectures and day-long meetings aimed at fostering coherence as well as cross-disciplinary contacts. A brief analysis of the articles shows the following distribution: "Exploration and History" 2; "Geology" 4; "Weather and Climate" 2; "Physical Oceanography" 6; "Chemistry" 5; "Phytoplankton, Primary Production and Microbiology" 5; "Zooplankton and Secondary Production" 16; "Fisheries" 10; and "Conflicting Uses" 8. Admirably, most authors set out lucidly the basics of their field before embarking on their application to Georges. Between the more serious contributions are engaging snippets with such titles as "How the Bank Got its Name" and "Bait Up: Dory Fishing on Georges Bank".

Few of the authors indulge in subject elitism. Rather, almost all of them seem acutely aware of the implications of their work for others interested in the Bank. The editorial policy persistently reinforces this unity. Thus the reader is constantly aware of recurrent themes: the influence of the last glaciation; proximity to the coast and its associated contaminants; canyons round the Bank; the ubiquitous effect of winter storms on the shallows; physical and biological seasonality; the contributions to work on the Bank from satellite remote sensing and submarine photography; the local 'mud patch'; the Bank's high biological productivity; difficulties of quantifying the energy flow through the trophic levels; difficulties of population measurement (and the inescapable urge of scientists to perceive and quantify biological diversity); and

comparison with other coastal regions. All this is set against the background of commercial fisheries and, more recently, exploration for oil.

The book is a paean to regional oceanography at its best. A history of Georges, it is also the history of marine science written small. The leitmotiv is a mood of sympathetic exploitation of the environment informed by a thoroughgoing catholic interest in the sea. Whatever your viewpoint — oceanography, fisheries, commerce or law — and whatever stage your career, you will profit both in your particular field and from the wide panorama of maritime endeavour that this book offers. □

J.H.S. Blaxter is Deputy Chief Scientific Officer and A. Edwards Principal Scientific Officer at the Dunstaffnage Marine Research Laboratory of the Scottish Marine Biological Association, Oban, Argyll PA34 4AD, UK.

Life under chilly conditions

F. Franks

The Effects of Low Temperatures on Biological Systems. Edited by B.W.W. Grout and G.J. Morris. Edward Arnold:1987. Pp.500. £45, \$89.50.

MULTI-author volumes tend to exhibit common strengths and weaknesses. Most chapters are written by experts in their field and are therefore authoritative. On the other hand, the strength of the chain is that of its weakest link. There is invariably at least one unreliable author who ignores deadlines, with the result that publication is delayed and the volume, when it finally appears, is not as topical as it could and should have been. This book bears the marks of such laggard contributors.

It is composed of 15 chapters, grouped into four sections: fundamental principles, techniques, environmental low-temperature biology, and applications. The chapters vary in length, detail, comprehensiveness and emphasis. Some authors have chosen to provide state-of-the-science overviews, whereas others critically discuss experimental evidence in order to draw conclusions. Yet others have limited themselves to summaries of practical freezing and thawing protocols. There is also a certain ambivalence in emphasis: is the book aimed at the student of the effects of cold on living organisms (that is, cold tolerance/resistance, acclimation) or at the cryopreserver (dimethyl sulphoxide, controlled cooling, liquid nitrogen)? If the former, then why is there no mention of microorganisms or insects? What determined the choice of the topics? The volume could have benefited from some ruthless editing.

For me, the three outstanding chapters are "Cells at Low Temperatures" (Morris and Clarke), "The Adaptation of Aquatic Animals to Low Temperatures" (Clarke) and "Low Temperature and Biological Electron Microscopy" (Skaer). Each is a masterpiece of description, critical analysis, conclusions and pointers for future developments; they are stylishly written, with useful bibliographies. Contributions dealing with higher plants (Wilson, Grout), mammalian hibernation (Wang), medical and clinical aspects (Fuller, Green) are also interesting, although not as incisive.

The point of a chapter on 'direct' chill injury (Morris) is not clear. The symptoms are brought about by fast cooling. In the ecosphere such cooling rates do not occur and in the laboratory one can presumably avoid them in order to achieve optimum results. Accounts of recent developments in optical cryomicroscopy (McGrath) and



National Geographic Society

Tree time — the picture, of resting chimpanzees, is reproduced from the new paperback edition of Jane Goodall's In the Shadow of Man, which was first published in 1971. The paperback has a brief introduction by Stephen Jay Gould, together with a very brief postscript by the author, and is published by Houghton Mifflin. Price is \$11.95.

the cryopreservation of parasites (James) will appeal to specialists. The chapter on plant germplasm preservation (Withers) reflects the underdeveloped state of that subject; at best it is recipe science. Withers makes a reasoned plea for *in vitro* plant germplasm preservation, but her bibliography, limited essentially to names of half-a-dozen senior authors, is only too revealing of the lack of importance attached to this subject by industry and governments, and hence by the scientific community. The final chapter on food freezing (Reid), eight pages in length, is no more than an appendix; one must wonder why it was written at all.

Despite a valiant attempt by the author to tackle a difficult and wide-ranging subject, the introductory chapter on physico-chemical principles (Taylor) poses problems. Some of the concepts (such as water structure, antemelting and incipient melting, bound water) are quite out of date. Some terminology is incorrect (for example, on p. 19 heat of crystallization is referred to as heat of fusion) and the material is not presented in the most effective sequence. In places Taylor side-

steps important, if complex issues, such as the freezing behaviour of ternary systems in which one component (glycerol?) does not crystallize. Instead, he deals at length with the familiar binary water-NaCl phase diagram. The reader is then referred to specialist articles which are not the easiest of reading.

The book is attractively produced and is essentially free from typographical errors, although some names are consistently misspelled. It is recommended reading for postgraduate students and others who wish to acquaint themselves with the elements of laboratory cryobiology and with the various responses of live cells and some organisms to low temperatures. □

F. Franks is in the Biopreservation Division, Pafra Ltd, 150 Cambridge Science Park, Cambridge CB4 4GG, UK.

Field surveys

Peter J. Smith

Geomagnetism, Vols 1 and 2. Edited by J. A. Jacobs. Academic: 1987. Vol. 1, pp.627, £62, \$112; Vol. 2, pp.579, £62, \$112.

GEOMAGNETISM, which may be broadly but incompletely defined as the study of the Earth's magnetic field, is a subject of curious contrasts. In academic circles it is often regarded as a highly esoteric topic seldom to be taught outside advanced geophysics courses; yet in the severely practical role of navigational aid it underpinned world exploration and trade for centuries. It has thus traditionally attracted an uneasy alliance of mathematical theoreticians and navy hydrographers. In its pure form as a study of the structure and origins of the Earth's magnetic field, it is often regarded by Earth scientists as an isolated topic unconnected with almost anything else; yet in a blaze of glory during the 1960s it gave rise to an offshoot, palaeomagnetism, that revolutionized the whole of geology. Though studied by few, it has thus influenced the thinking of every Earth scientist in the world.

There are internal contrasts, too, for the subject comprises two fairly distinct fields, generally studied by two quite different groups of people. When Chapman and Bartels published their huge two-volume work *Geomagnetism* in 1940, 'geomagnetism' was revealed as largely the study of short-term variations in the small part of the Earth's magnetic field generated externally — a branch of solar and atmospheric physics, in fact. The main (internal) geomagnetic field could be described very quickly, and there was nothing at all to be said about its causes. Little had apparently changed in a relative sense when the two volumes of Matsushita and Campbell's multi-authored *Physics*

of *Geomagnetic Phenomena* appeared in 1967. Solid-Earth geomagnetism had expanded, but then so had its external counterpart, aeronomy, largely as a result of the advent of rockets and satellites.

If the two branches of the subject *must* be dealt with together, there is now clearly a case for giving solid-Earth geomagnetism a fairer share of the proceedings; and that is what Jacobs evidently intends that *Geomagnetism*, a new multi-authored work, should do. It was envisaged that the two volumes published so far would completely cover the solid-Earth aspects of geomagnetism and that they would be followed (in about a year's time) by two further volumes devoted exclusively to aeronomy — a convenient split. Unfortunately, for reasons too complex to be explained here, some of the solid-Earth topics have had to be deferred. Those interested solely in the main field will therefore later have to beg, borrow or steal at least one other volume containing much material of little concern to them; but they will hardly be able to complain that the balance of topics has not been handsomely redressed in their favour.

In the first two volumes, then, ten specialists cover the history of geomagnetic observation, instrumentation in general, instrumentation and experimental methods for oceanic observation, the main geomagnetic field and its changes, the crustal field, magnetohydrodynamics of the Earth's core, the dynamics and kinematics of the origin of the main field (two separate chapters), lunar palaeomagnetism, planetary magnetism, and the magnetism of meteorites. The topics to come are magnetotellurics, terrestrial palaeomagnetism and lunar magnetism. Although etymological purists may object to the trend, it is perhaps worth emphasizing that 'geomagnetism' now includes the magnetism of extraterrestrial bodies. Indeed, over the past two decades or so, beyond-Earth studies have been the subject's chief growth area.

Because any one of the above-mentioned topics could make a two-volume set in itself, comprehensive treatment within a single chapter is hardly to be expected, although some chapters (especially that on the main field) manage to pack in a surprising amount of detail. Nevertheless, all of the contributions provide good, up-to-date general surveys of their topics (albeit only for the mathematically proficient) and will thus prove useful starting points for the new postgraduate at whom they are presumably aimed. Chapman and Bartels were able to produce a detailed treatise on all existing knowledge of the subject. Forty-eight years of work later, such luxury is regrettably no longer feasible, even with a large team of writers. □

Peter J. Smith is a Reader in the Department of Earth Sciences, Open University, Walton Hall, Milton Keynes MK7 6AA, UK.

Mutating molecules

N.H. Barton and J.S. Jones

Molecular Evolutionary Genetics. By Masatoshi Nei. *Columbia University Press*: 1987. Pp. 512. \$50, £29.50.

In population genetics, the theoretical tail has usually wagged the experimental dog. At a time when mathematicians were producing elegant models of evolutionary change, biologists were still struggling with problems which seemed insoluble: how much genetic variation is there, what is the strength of natural selection and what genetic changes are involved in the formation of new species?

However, the startling new discoveries of molecular biology — homology between distantly related organisms, quantities of apparently functionless DNA, widespread gene duplication which may lead to volcanic eruptions of pseudogenes from those which still retain a function, and the concerted differentiation of repeated sequences — all now demand explanation by population genetics theory. Even the discrete nature of the gene looks increasingly foggy, and pea soup, rather than peas, is now the geneticist's staple diet. As Nei points out, there is an urgent need for molecular biology and population genetics to be unified into 'molecular evolutionary genetics'. His book attempts to do that, and provides a valuable compendium of up-to-date molecular information and of the statistical methods by which it can be described.

After an account of the evolutionary history of life and of the remarkably good general fit between the amino acid sequences of particular proteins in different species and their dates of divergence, Nei goes on to the mass of new data on the evolution of nucleotide sequences. Just as with proteins, rates of evolution at the DNA level seem to depend greatly on 'functional constraints'. Some sequences have scarcely changed during the history of life, presumably because purifying selection removes any mutations which deviate from the functional norm. In other cases, however, related species may diverge very rapidly in DNA sequence, a pattern usually interpreted as a result of the relaxation of selection when no functional transcript is produced.

This reflects a remarkable difference in emphasis between molecular and morphological evolution. Before the advent of molecular biology, rapid change was seen as a proof of natural selection, whereas it is now generally taken as evidence against it. In the same way, molecular evolution has it that nature abhors a polymorphism; genetic variation once thought to demand an explanation in

terms of balancing selection is now seen simply as the passage of a transient neutral variant through a population. As is so often true in evolutionary biology, these apparent contradictions may be due largely to differences in the timescales involved: neutral sequence divergence takes place over millions of years, whereas selection can cause drastic phenotypic changes in a few generations.

Nei goes on to give an account of a variety of sequence alignment and tree-building methods which can introduce molecular biologists to evolutionary data that they never knew they had. Molecular information is invaluable in providing an objective basis for the phylogenetic classifications which are at the centre of much of biology. He ends with a personal view of the 'new evolutionary biology' that may be needed to cope with modern molecular genetics. Somewhat surprisingly, this turns out to be close to the theories of evolutionary genetics which first emerged from the rediscovery of Mendelism at the turn of the century. Nei argues that mutation is the creative process at both the molecular and the phenotypic level: natural selection is primarily conservative, and simply weeds out the unfit. His point that "natural selection does not create any new genotypes" is reminiscent of arguments such as that of Johanssen in 1915 that "selection of differing individuals creates nothing new". This emphasis on the importance of mutation can be misleading: selection and mutation work together in a long process of trial and error which can build up adapted genotypes that would stand almost no chance of assembly by random mutation alone.

It is particularly hard to see how mutation could shape adaptations if these involve the gradual accumulation of many genetic changes of small individual effect. Nei argues that most adaptations involve just a few major mutations. Although the drastic morphological changes induced by homoeotic and heterochronic mutations support this idea, the discovery that gene regulation can depend on a diffuse set of sequences, well outside the coding region, points in the opposite direction.

Nei's bold conclusions sit slightly oddly with the inevitable uncertainties involved in explaining the torrent of new molecular information: indeed, it may take as long to reach a proper understanding of the natural history of molecules as it has to elucidate that of whole organisms. Nei's book gives a remarkably complete and up-to-date account of the present state of the relationship between theory and practice in molecular evolution. But we should not be surprised if his second edition turns out to be very different from the first. □

N.H. Barton and J.S. Jones are in the Department of Genetics and Biometry, University College London, 4 Stephenson Way, London NW1 2HE, UK.

In the words of the good and the great

A.P. French

The World of Physics: A Small Library of the Literature of Physics from Antiquity to the Present, Vols I-III. By Jefferson Hane Weaver with commentaries and notes by Lloyd Motz and Dale McAdoo. *Simon & Schuster: 1987-1988. Each volume about 1,000 pages, \$29.95, £25.*

As THE compiler of this far-from-small collection says in his preface, "An anthology is a work of prejudice". So too, of course, is any review of such an enterprise: one is bound to measure its success in terms of one's own ideas of what should or should not have been included. By any standard, however, this set of volumes is an impressive accomplishment, both for its sheer magnitude and for the care with which it has been put together and edited.

The subtitles of the separate volumes that make up the work are I: *The Aristotelian Cosmos and the Newtonian System*; II: *The Einstein Universe and the Bohr Atom*; III: *The Evolutionary Cosmos and the Limits of Science*. These titles in themselves are indicative of two things: the

scope of the collection goes beyond what would normally be understood by the word 'physics', and there is a very strong emphasis on the twentieth century. The anthology is explicitly described as "a library of the literature of physics from antiquity to the present", and it seems clear that the intention has been to depict the historical growth of the subject, not just to provide a survey of our present knowledge. But even before exploring the books in detail, one is bound to wonder if justice has been done to physics before 1900. This doubt is reinforced when one looks at the actual contents of Vol. I, and my chief criticism of the anthology is concentrated in this area.

Volume I begins enticingly with a section entitled "What is Physics?", containing essays bearing on this theme by four great physicists — Feynman, Born, Einstein and Planck. But the next 300 pages are devoted to what at best can be described as pre-physics, ranging from ancient mythology and astronomy up to Copernicus. Appropriate extracts from Aristotle and Lucretius are included, but so, too, are various poetical, mystical or simply mystifying excerpts from such sources as the Bhagavadgita, Maimonides and Paracelsus. These are interesting in themselves, perhaps, but of negligible relevance to the growth of physical thought.

What is troublesome is not so much the inclusion of this material, but the fact that all of classical physics, from Galileo to Maxwell, is then compressed within the remaining half of the first volume. Particularly hard to understand are the cavalier treatment of optics (limited to brief excerpts from Huygens and Young) and even more the dismissal of electromagnetism through a ten-page extract from Faraday's *Experimental Researches* (his discovery of induction, of course) and a similarly brief excerpt from Maxwell's "Dynamical Theory of the Electromagnetic Field".

It is a relief to turn to Vol. II, which contains a splendid presentation of topics in twentieth-century physics through the words of many of its principal contributors. Almost half of the 60 or so pieces in this volume are excerpts from Nobel lectures, providing a vivid picture of the development of radioactivity, relativity, quantum theory and particle physics. There is even an excursion into the contributions of physics to fundamental biology, with passages from such figures as Schrödinger, Crick and Delbrück. (It was, however, surprising to find no selection devoted to the discovery of X rays, surely one of the most exciting and influential events at the beginning of the modern era.)

The third volume begins, rather arbitrarily, with a section on mathematics; perhaps the most interesting of the four

selections here is Eugene Wigner's famous essay "The Unreasonable Effectiveness of Mathematics in the Natural Sciences". But most of Vol. III is devoted to astrophysics and modern cosmology, and to a collection of 18 articles concerned with the relation of physics and philosophy. Included in this volume are sections on black holes, on theories of the origin of the Universe and on unification theories of the forces of nature. Here are to be found excerpts from three more Nobel lectures, by Weinberg, Salam and Glashow. Many other distinguished names are also represented, including Bethe, von Weizsäcker, Hubble, Bondi, Weisskopf, Eddington, Wheeler and Hawking.

The compiler says in his preface that, in making his selections, he had in mind the educated novice, not the specialist, and to this end he chose writings that were mostly free of equations and technical jargon. His aim is a laudable one, but I think that there are many cases where the mere absence of formalism does not make the subject matter easily accessible.

Very few people, I suspect, will have the stamina to work systematically through this collection, but there is certainly a great deal to reward a more selective approach. The compiler states that his aim was "to strike a pleasing balance between an encyclopedia, a sourcebook, a textbook and a history". I rather question the textbook role, but I think that there is some validity to the claim for the other three. Almost every individual selection is preceded by short introductory comments — often biographical, and generally helpful and informative (though those about energy and entropy are extraordinarily confusing). The anthology is further enhanced by an abundance of apposite quotations heading the individual selections.

I must register one general criticism. The title, *The World of Physics*, is not really justified. Quite apart from the inexcusably short shrift given to the nineteenth century, the emphasis with respect to recent developments is heavily on elementary particles and cosmology, which engage only a small fraction of the world's physicists today. One looks in vain for any significant discussion of nuclear structure, modern optics (including lasers) or the physics of condensed matter (except for Bardeen and Shockley on the transistor). Overwhelmingly, too, the authors quoted are theorists. This is, indeed, predominantly an anthology about fundamental theory. Within that limitation, however, it is a praiseworthy production, for which the self-effacing compiler (about whom we are told nothing) is to be commended. □

A.P. French is a Professor in the Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

ADVERTISEMENT



The ANS Journals

Your number one source for current information about

NUCLEAR ENERGY

- NUCLEAR NEWS
- NUCLEAR SCIENCE & ENGINEERING
- NUCLEAR TECHNOLOGY
- TRANSACTIONS
- REMOTE SYSTEMS TECHNOLOGY PROCEEDINGS
- FUSION TECHNOLOGY

For free catalog contact:

AMERICAN NUCLEAR SOCIETY
 555 N. Kensington Avenue
 La Grange Park, Illinois 60525 USA
 312-352-6611 Telex: 4972673

Diffuse back-arc deformation in the southwestern Pacific

Michael W. Hamburger* & Bryan L. Isacks†

* Department of Geology, Indiana University, Bloomington, Indiana 47405, USA

† Department of Geological Sciences, Cornell University, Ithaca, New York 14853, USA

Earthquake distribution and focal mechanisms from the Lau and North Fiji back-arc basins indicate a diffuse and shear-dominated system of deformation, at odds with traditional plate tectonic models of lithospheric accretion in back-arc basins. The entire back-arc region between the Tonga and New Hebrides arcs is more realistically modelled as a giant, internally deformed, pull-apart basin, formed along a left step in the transform boundary between the Pacific and Indo-Australian plates.

THE back-arc basins of the western Pacific, located adjacent to active compressional plate boundaries, have long been recognized as sites of extensional deformation. Although the hypothesis of recent extension in these basins has been generally accepted, the cause and the mechanism of formation have not been satisfactorily explained. Despite strong evidence that active extension is taking place, it remains unresolved whether the mechanism of back-arc deformation is analogous to that observed along the global system of mid-ocean ridges.

A series of fundamental geological and geophysical characteristics link the back-arc basin and the mid-ocean ridge tectonic environments as zones of active lithospheric accretion¹⁻⁴. Both are characterized by shallow sea-floor depths with high heat flow and thin sediment cover. They share anomalously low mantle velocities and high seismic attenuation, active crustal seismicity, and thin oceanic crust. Basement rocks in both environments are commonly composed of tholeiitic basalt. These common characteristics have led to the obvious tectonic analogy with sea-floor spreading at the mid-ocean ridges. The back-arc basins' spreading centres, although often identifiable, are, however, less clearly defined morphologically and geophysically than the mid-ocean ridges^{5,6}. Many investigators have nonetheless contended that lithospheric accretion in the back-arc basins usually takes place along discrete spreading centres similar to those observed in the major ocean basins^{3,4}.

An alternative model^{6,7} holds that extension in many marginal basins takes place by a more diffuse system of back-arc extension. Lithospheric generation in these environments often appears to be complicated by the presence of point-source volcanism (seamounts), short-lived spreading ridges, and rapid jumping of ridge crests. Two of the best studied back-arc basins of the southwestern Pacific are examined here as a unique testing ground for these conflicting models. Because of their high seismicity and the clear spatial separation of plate boundary and back-arc seismic activity (active seismic zones in the North Fiji and Lau basins are separated from the Tonga and New Hebrides plate boundaries by as much as 800 km), these two basins are arguably the ideal test areas for the conflicting theories of back-arc deformation.

Tectonic setting

The back-arc deformation in the southwestern Pacific must be considered in the context of plate convergence between the Pacific and Indo-Australian plates (Fig. 1). The portion of the plate boundary located between the Tonga and New Hebrides arcs, henceforth referred to as the interarc region, is unusual in several important respects: (1) the plate boundary is discontinuous, jumping from westward subduction at northern Tonga to eastward subduction at the southern New Hebrides; (2) it is one portion of the convergent plate boundary that is not domi-

nated by subduction tectonics; (3) a broad and complex zone of Cenozoic tectonism separates the active trenches; (4) two east-trending bathymetric deeps, the Vitiaz Trench and the Hunter Fracture Zone (Fig. 2a), mark the sites of fossil subduction zones^{8,9}. In the most synoptic view, the entire interarc zone can be considered as a complex transform zone linking the two opposite-facing trenches.

A major active component of the Pacific/Indo-Australian boundary is the Fiji Fracture Zone (FFZ). The FFZ is most evident as a continuous belt of seismicity and rough topography, extending westward from the northern termination of the Tonga Trench (Fig. 2a). Seismicity and strike-slip focal mechanisms confirm its identification as the southern, transcurrent boundary of the Pacific Plate in this area^{8,10-12}.

Independently of direct evidence for back-arc deformation, a strong circumstantial argument for active extension in the interarc region originates from the geometric arrangement of plates. The Tonga and New Hebrides trenches are of opposite polarity; in the absence of intervening subduction zones, the transform zone connecting these opposing trenches must lengthen at a rate at least equal to the plate-convergence rate¹³. Given the east-west plate-convergence direction¹⁴ and the considerable latitudinal overlap of the two trenches, such divergence cannot occur by simple expansion of a trench-trench transform fault. The divergence of the two trenches occurs largely by extensional deformation in the North Fiji and Lau basins.

Direct evidence of active extension in the North Fiji and Lau basins is provided by their morphology, thin sediment cover, seismicity, heat flow, and the presence of young tholeiitic basalts, similar to those erupted at mid-ocean ridge spreading centres^{1,8,15-17}. Pervasive anomalous upper mantle velocities^{18,19} and unusually high attenuation of seismic waves²⁰ support the contention of asthenospheric upwelling beneath the basins. Chase⁸, in the first detailed analysis of back-arc deformation in the interarc region, defined stable microplates and active microplate boundaries to explain the active tectonism in the area; spreading centres were defined largely by inference from the distribution of shallow earthquakes and poorly defined marine magnetic anomalies. The proposal⁸ of a major north-trending spreading centre in the central portion of the North Fiji Basin (between 173° and 174° E) has largely been supported by more recent marine geophysical studies^{9,21,22}. But the spreading centre that was identified by Chase⁸ is not the sole locus of sea-floor spreading; other active spreading ridges, with varying positions and orientations, have been documented elsewhere in the North Fiji Basin²²⁻²⁶.

Similarly, a complex array of sea-floor spreading configurations has been proposed for the Lau Basin. Various investigators have proposed that spreading centres in the Lau Basin are oriented northwest-southeast⁸, northeast-southwest^{15,16} and

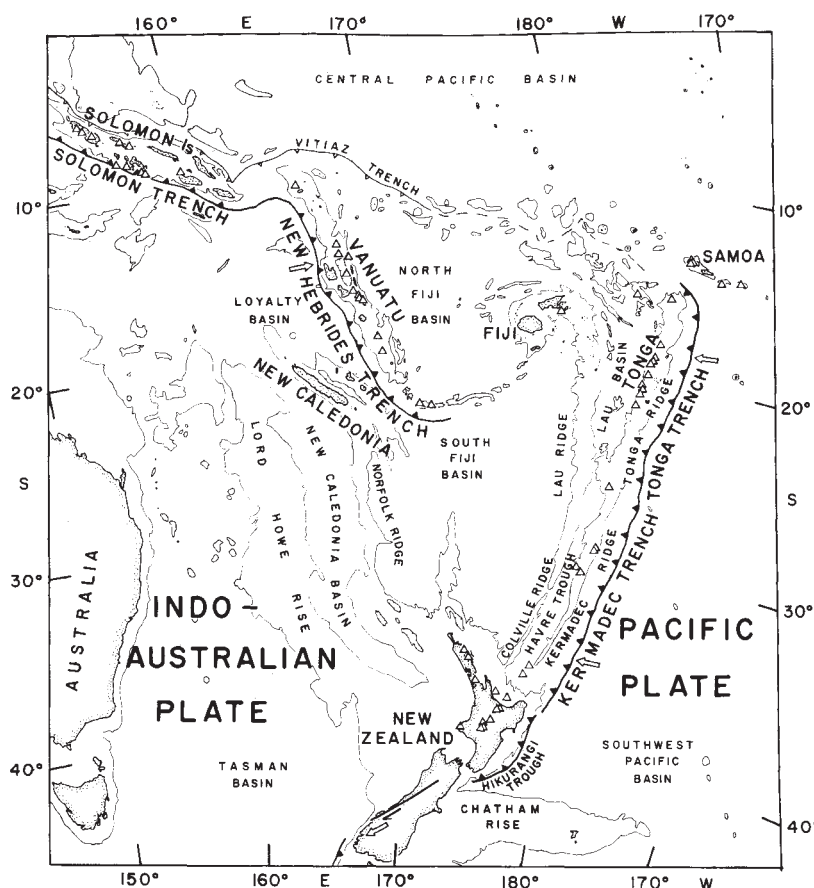


Fig. 1 Tectonic and morphological features of the Pacific/Indo-Australian plate boundary. Open arrows indicate direction of relative plate convergence. Contour line shows 2-km isobath. Holocene volcanoes are indicated by open triangles. Data on bathymetry, volcanoes and plate motions are from ref. 49.

north-south^{9,27}. Recent marine geophysical investigations^{26,28} have confirmed the presence of well-defined active spreading centres over short distances (10–20 km) in the central and southern Lau Basin. The presence of widespread seamount volcanism¹⁶, and the lack of well-defined linear magnetic anomalies correlatable over distances greater than 30–50 km (ref. 5) suggest, however, that a mode of disorganized extension, characterized by abundant short-lived spreading ridges and point-source volcanism, could predominate in the Lau Basin.

Distribution of shallow earthquakes

To investigate the seismic manifestation of active back-arc deformation, we have compiled the best available hypocentral information for teleseismically recorded shallow earthquakes (magnitudes greater than ~4.5) located between the Tonga and New Hebrides arcs. The earthquake catalogue is based on contributions from eight previous compilations. Criteria for inclusion in the catalogue, as well as complete hypocentral data, are presented elsewhere²⁹.

Epicentres for the earthquakes located in the interarc region are presented in Fig. 2a. The data for the interarc region (within the dashed lines) include 727 earthquakes, compared to only 260 events used in the previous comprehensive seismicity study³⁰. Furthermore, we have discarded poorly located events (such as those located by <15 stations or assigned 'C' quality) which were included in previous compilations, thus significantly improving the reliability of the epicentral distribution reported here.

The location errors for events in the interarc region vary considerably and are largely a function of the number and azimuthal distribution of recording stations. For the smallest events, recorded by fewer than 15 stations, gross mislocations of earthquakes can occur; these events were thus not considered. As the epicentral reliability increases with the number of record-

ing stations, well-located events, recorded by 40 or more stations, are expected to be within 10–15 km of the teleseismic location.

Seismic activity in the interarc region is not limited to linear seismic belts as previously suggested^{8,30}. Seismic deformation within the basins is far more broadly distributed than implied by microplate tectonic models. A concentration of seismicity within an ill-defined, north-trending zone in the central North Fiji Basin, between 172° and 174° E, is suggested. This belt has been linked to extension along the major North Fiji Basin spreading centre^{8,9,21}, but the breadth and irregularity of the seismic zone are in sharp contrast to the narrow seismicity belts associated with the mid-ocean ridge system. Well-located events in this zone are distributed at distances up to 100 km away from the simple meridional spreading ridge. A second area of high activity, immediately west of the Fiji Platform (around 176° E), has also been proposed as a site of possible active spreading^{8,23}. Observations of detailed submarine morphology²¹ and earthquake focal mechanisms¹², however, suggest a more complex zone of distributed shear deformation, with localized zones of extension.

Lau Basin seismicity is dominated by two main features: the Fiji Fracture Zone, which bounds the basin in the north, and Peggy Ridge, a northwest-trending linear bathymetric high immediately east of the Fiji Platform. The overall seismicity trend and earthquake focal mechanisms clearly define the FFZ as an active transcurrent boundary of the Pacific Plate; however, the breadth of the zone, exceeding 200 km in the east, suggests a plate boundary zone more complex than just a simple transform fault. Events associated with the Peggy Ridge define a clear northwest-trending lineation, parallel to the bathymetric feature and near-vertical dextral fault planes²⁹. Other areas of high activity in the central and southern Lau Basin may be associated with active spreading centres. But well-located shallow earthquakes in the central Lau Basin (18°–19° S) are dis-

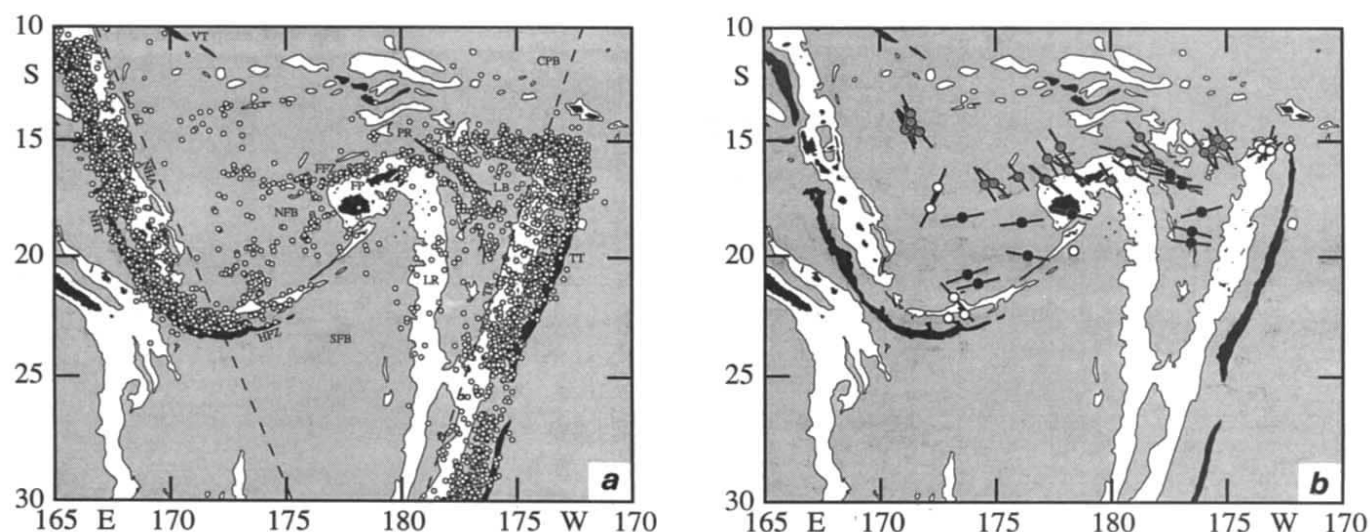


Fig. 2 Earthquake distribution and focal mechanism solutions in the interarc region of the southwestern Pacific. *a*, Compilation of well-located shallow events ($h \leq 70$ km) is shown for the interarc region (within dashed lines); events along the Tonga and New Hebrides plate boundaries, located by the United States Geological Survey, are shown for comparison. Sea-floor morphology of the region³⁵: light shading indicates areas >2 km depth; heavy shading shows areas >7.5 km for Tonga Trench, >5 km for New Hebrides and Vitiaz trenches. Bathymetric features discussed in text: CPB, Central Pacific Basin; FFZ, Fiji Fracture Zone; FP, Fiji Platform; HFZ, Hunter Fracture Zone; LB, Lau Basin; LR, Lau Ridge; NFB, North Fiji Basin; NHA, New Hebrides Arc; NHT, New Hebrides Trench; PR, Peggy Ridge; SFB, South Fiji Basin; TR, Tonga Ridge; TT, Tonga Trench; VT, Vitiaz Trench. *b*, Summary of focal mechanism types. Earthquake types are defined by consistent orientations of principal stress axes. Heavy bars indicate horizontal projections of mechanism tension (T) axes. Type-1 events (shaded circles) are defined by NE-trending compression (P) axes, NW-trending tension (T) axes and nearly E-W fault planes; type-2 events (filled circles) are defined by E-W T axes, N-S P axes, and NW- and NE-striking fault planes. Other mechanism types are shown by open circles.

tributed over 100 km away from the spreading ridge identified in the central Lau Basin^{26,27}.

Earthquake focal mechanisms

As part of this study, we compiled all the reliably determined focal mechanisms for 51 shallow earthquakes in the interarc region, including seventeen events obtained by Harvard/PDE centroid-moment tensor inversion³¹. The mechanism parameters and detailed interpretation are summarized elsewhere^{12,29}. There is significant diversity among the focal mechanisms, but several first-order patterns are evident. Most significantly, this extensional tectonic regime is dominated by strike-slip deformation. With the exception of earthquakes located at the complex terminations of the Tonga and New Hebrides arcs and three anomalous events within the interarc zone, all the larger events in this region (namely, those for which teleseismic mechanisms could be determined) occur by strike-slip faulting. A second important observation is that the stress orientations of earthquakes within the interarc region are inconsistent with the

east-west normal faulting or transform faulting earthquakes that are predicted by the mid-ocean ridge tectonic models³².

With few exceptions, the mechanisms in the interarc region fall into two well-defined groups, summarized in Figs 2*b* and 3: type-1 earthquakes (shaded circles in Fig. 2*b*), located close to and north of the Fiji Fracture Zone, are characterized by northeast-trending compression (P) axes, northwest-trending tension (T) axes and nearly east-west (sinistral) fault planes. Type-2 earthquakes (filled circles in Fig. 2*b*), located south of the FFZ, are characterized by east-west T axes, north-south P axes, and northwest- and northeast-striking fault planes.

Both fault plane orientations can be interpreted in terms of the relative motion between the Pacific and Indo-Australian plates. Type-1 mechanisms occur along the Fiji Fracture Zone, at the boundary of the Pacific Plate. Their east-trending fault planes accommodate the westward translation of Pacific Plate lithosphere along a transform plate boundary. Type-2 mechanisms occur only in the zone of overlap of the Tonga and New Hebrides trenches. They indicate a zone of pervasive shear deformation; their east-trending T axes reflect the sense of dominant extension in the interarc region. Although there is an inherent ambiguity in the inference of principal stress orientations from earthquake focal mechanisms³³, we regard the large and systematic difference between the two geographically isolated groups of mechanisms as reflecting underlying patterns in the actual regional stress field. Nonetheless, the variability in inferred stress orientations for individual events within each group (Fig. 3) can be partly explained by the presence of pre-existing zones of weakness.

The exceptions to this pattern occur in discrete geographic and tectonic zones. A large group of events at the northern termination of the Tonga subduction zone reflects the extraordinary complexity of this area. Subduction of the Pacific Plate (beneath northernmost Tonga) gives way to tearing of the plate and westward translation (by transcurrent faulting) along the FFZ. The mechanisms in this area indicate dip-slip faulting along near-vertical, east-trending fault planes and reverse faulting mechanisms within the Pacific Plate^{10,29}. Earthquakes at the

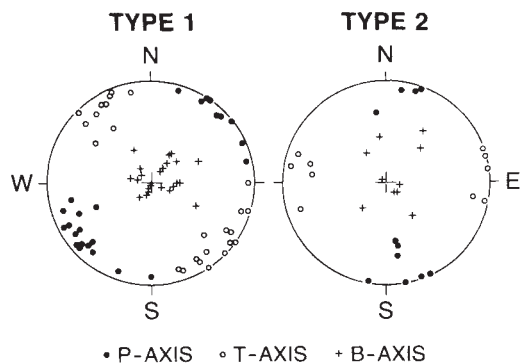


Fig. 3 Inferred principal stress axes (P , compression; T , tension; B , null) for type-1 and type-2 focal mechanisms. Locations are shown in Fig. 2*b*.

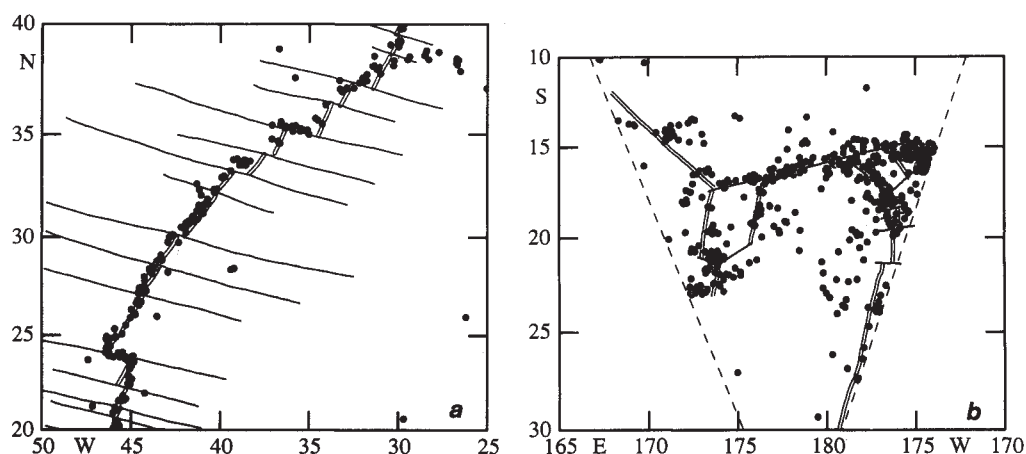


Fig. 4 Comparison of mid-ocean ridge seismicity with that of the interarc region. Shallow earthquakes ($h \leq 70$ km), located by ≥ 20 stations, are from the International Seismological Centre catalogue, 1964–1984. *a*, earthquakes along the mid-Atlantic Ridge; ridge segments (double lines) and fracture zones (single lines) are from ref. 50; *b*, Earthquakes in the interarc region, exclusive of the Tonga and New Hebrides plate boundaries. Proposed microplate boundaries, as in *a*, are from ref. 49. Note that the scale of the two maps is roughly the same.

southernmost edge of the North Fiji Basin indicate similar vertical dip-slip faulting of the Indo-Australian Plate¹². Their east-trending fault planes permit the northern portion of the plate to be subducted at the New Hebrides Trench, while the southern portion translates eastward at the Earth's surface.

A pair of solutions for events situated within the active belt of seismicity between 172° and 174° E (open circles in Fig. 2*b*) provide the only direct seismic evidence of extensional deformation within the North Fiji Basin. The stress directions inferred for these two events are virtually identical, but are strikingly dissimilar to those from the remainder of the basin. They clearly deviate from the east-west extension direction predicted by the major north-striking spreading ridge^{8,9,21}. These events are more easily explained by their proximity to the western termination of the Fiji Fracture Zone (Fig. 2*a*). Such fault terminations are widely recognized as areas of unusual stress concentration³⁴. Thus, the unusual stress orientation could represent a local deviation of the regional stress field associated with the divergent flank of a strike-slip fault termination.

Discussion

Extrapolation of the mid-ocean ridge model of sea-floor spreading to the back-arc environment can simply explain the kinematics of back-arc basin development, but the detailed geophysical observations from the interarc region present a series of significant contradictions to the mid-ocean ridge spreading paradigm:

- (1) The broadly distributed shallow seismicity and seismic attenuation in the interarc region indicate that deformation is not limited to narrow belts of activity associated with spreading centres and transform faults. There is a fundamental contrast between mid-ocean ridge seismicity and that of the interarc region. Figure 4 compares two extensional plate boundary zones of nearly identical scale: the northern mid-Atlantic Ridge and the interarc region. Whereas mid-Atlantic Ridge activity is confined to a narrow, linear belt that can be traced along well-defined transform-fault offsets, the interarc seismicity is altogether diffuse, with well-located events positioned far from identifiable spreading centres.
- (2) The focal mechanisms observed in the central Lau and North Fiji basins are dominated by northwest- and northeast-trending nodal planes. These fault plane orientations, even if they are interpreted as transform faulting associated with active spreading ridges, cannot be reconciled with the proposed north-trending spreading centres in the basins.
- (3) Active volcanism has been demonstrated at widely separated localities throughout the Lau and North Fiji basins. At least eight different zones of active extrusion have been identified in the two basins^{16,21–26}.
- (4) The simple rectilinear mid-ocean ridge/transform system predicted by the interpretation of magnetic anomaly data in the

North Fiji Basin is not backed up by the sea-floor morphology. The regional bathymetric fabric³⁵, as well as the detailed bathymetry from two minor spreading ridges in the central North Fiji Basin^{24,25}, show clear northeasterly trends.

(5) The analyses of magnetic anomaly data, on which the spreading histories are based, are themselves suspect. Lawver *et al.*⁵ showed that for the Lau Basin, in areas of dense data coverage, the magnetic anomalies are highly irregular and possible to correlate only over short distances. There is evidence for a broad axial magnetic high associated with the central North Fiji Basin spreading centre^{8,9,21}, but the dating of the basin crust by correlation of the flanking magnetic anomalies with synthetic profiles is unreliable at best. Such irregular magnetic anomalies could be explained by development of short-lived spreading centres, rapid jumping of ridge crests, and seamount volcanism distributed over a broad axial zone⁶.

(6) The virtual absence of normal faulting contrasts with that observed along many mid-ocean ridges; given the amount of extension required in the interarc region, the paucity of normal faulting focal mechanisms demands that the marginal basin lithosphere be exceptionally thin, weak, and unable to support large stresses. This state of the lithosphere is consistent with that proposed to result from broad, diffuse zones of upwelling, random locations of magma leaks, and slow cooling of the back-arc lithosphere⁶.

(7) In virtually all traditional plate tectonic maps of the south-western Pacific, the Fiji Platform (Fig. 2*a*) is treated as a stable microplate surrounded by zones of active tectonism. But our detailed studies on the structure and deformation of the platform^{11,29,36} demonstrate that it is clearly distinguished from stable microplates: (i) the thin crust (15–20 km) and low mantle velocities beneath the platform ($V_p \sim 7.6$ km s⁻¹) indicate a thermally attenuated lithosphere lacking the thick, rigid 'mantle lid' that underlies stable continental and oceanic crust; (ii) the presence of Neogene volcanism and faulting³⁷, as well as widespread seismicity, within the Fiji Platform indicates a degree of tectonic activity that is not characteristic of stable microplates; (iii) earthquake focal mechanisms within the platform, similar to those in the neighbouring back-arc basins, indicate that the platform is subject to the same stress system of east-west extension that pervades the entire interarc zone.

(8) The rapid Neogene rotations of the Tonga, Fiji, and New Hebrides arcs^{38,39} require a continuously changing geometry of the interarc region, and a continuously evolving configuration of spreading centres. This situation is not geometrically compatible with the development of long-lived linear spreading configurations.

The widespread distribution of volcanism, seismicity and deformation in the interarc region cannot be reconciled with the predictions of a mid-ocean ridge model for back-arc extension; a more plausible model of back-arc deformation is

presented in Fig. 5. In this model, deformation is broadly distributed throughout the interarc region. Extension takes place not along linear spreading centres, but by pervasive shearing and localized crustal generation at irregular, short-lived spreading centres.

At a regional scale, the entire interarc region can be considered as a zone of 'transtension'³⁴, in which the dominantly transcurrent motion of the Pacific and Indo-Australian plates is accompanied by a significant component of extension. Extensional structures are observed along most strike-slip faults, ranging from a laboratory or outcrop scale to a regional scale^{34,40-42}. A well-documented cause for localized extension along strike-slip faults is 'releasing bends', where the main locus of wrench faulting takes a lateral step. Such conditions can result in the formation of major extensional 'pull-apart basins'⁴⁰. Within such basins, strike-slip faulting is commonly observed, along synthetic or conjugate 'Riedel shears', as well as normal slip-on faults oriented obliquely to the main bounding faults³⁴.

In this framework, the interarc region can be viewed as a giant (700 km) left step between two major transcurrent faults, the Fiji Fracture Zone and Hunter Fracture Zone. This step offsets the westward motion of the Pacific Plate (north of Tonga) from the eastward motion of the Indo-Australian Plate (south of the New Hebrides). Within the pull-apart basin a large amount of strain is accommodated by strike-slip faulting, along synthetic and conjugate Riedel shears. The northeast-oriented sinistral strike-slip faults within the North Fiji Basin fit the orientation and sense of motion predicted for the Riedel shears; the regional bathymetric fabric in the basin corroborates this fracture pattern. In two cases within and near the Fiji Platform, the northeast-trending fault planes of earthquake focal mechanisms could be identified as the fault planes, based on ancillary geological and seismological evidence²⁹.

Faulting in the northern Lau Basin is dominated by strike-slip faults oriented perpendicular to the North Fiji Basin fault trends, as suggested by the well-defined northwesterly bathymetric fabric and seismicity trends. These conjugate shears are less commonly observed in model experiments, but are important components of many wrench fault systems observed in the field⁴², and could be the result of fault blocks rotated by movements between major wrench faults. In the southern Lau Basin, northeast-trending fabric, like that observed in the North Fiji Basin, predominates. Approaching the Fiji Fracture Zone, the fabric appears to gradually rotate toward the FFZ. Thus, the same stress system is accommodated along northeast-striking fault planes in the southern Lau Basin, and along northwest-striking fault planes in the northern portion of the basin.

The diffuse back-arc deformation documented in the North Fiji and Lau Basins may represent an end-member in the range of possible modes of back-arc deformation. Detailed observations from the Scotia Sea⁴³, Bismarck Sea⁴⁴, and the Mariana Trough⁴⁵ suggest spreading about discrete central rift zones. In these cases, back-arc seismicity appears to be isolated to the axial rift and transform zones^{44,46}. Several inactive marginal basins apparently record similar mid-ocean ridge-type spreading histories^{3,4}. In contrast, the evolution of several other relict basins, notably the Sea of Japan^{7,47} and the Tyrrhenian Basin⁴⁸, suggest basin development by diffuse spreading, with multiple rift centres, similar to the present-day North Fiji and Lau Basins.

For each of these diffuse-spreading basins, a case can be made that the spreading history is a result of unusual tectonic settings. Rapidly evolving configurations of major plate boundaries may in turn produce rapidly evolving basin geometries. In the case of the interarc region, the geometry of the basin is not determined by some preexisting configuration of the basin's bounding faults; rather, a rapidly changing basin geometry is ordained by the peculiar clockwise rotation of the New Hebrides arc³⁸. Such evolving boundary conditions, rather than subduction zone dip angle⁷, may be the primary control on the mode of back-arc spreading.

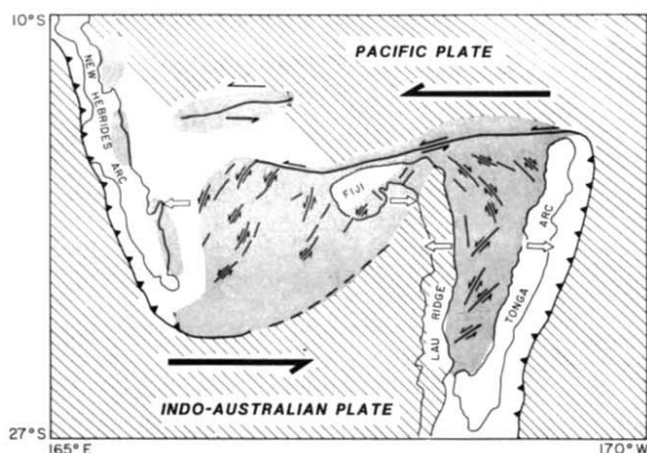


Fig. 5 Schematic model of deformation in the interarc region. Tectonism in this area is dominated by relative motion between the Pacific and Indo-Australian plates, indicated by heavy arrows. Within the area of overlap of the Tonga and New Hebrides trenches, motion between the major plates results in extensional deformation of the interarc zone, as a 'pull-apart basin' (indicated by open arrows), bounded by offset sinistral shear faults (the Fiji and Hunter fracture zones). Stable areas, within the major plates, are shown by ruling. Areas of known active deformation within the interarc zone are indicated by shading. Unshaded areas may or may not be actively deforming. Fracturing within the North Fiji Basin and southern Lau Basin is characterized by shear faulting along northeast-trending fault planes, and within the northern Lau Basin by antithetic shearing along northwest-trending fault planes.

We are grateful to D. Sarewitz and A. Bloom for discussions and to M. Barazangi, J. Brophy and three anonymous reviewers for reviewing this paper. Figures were drafted by E. Farkas and the drafting sections of the Fiji Mineral Resources Department and the Indiana Geological Survey. The manuscript was typed by J. Isaacs and D. Still. This work was supported by the United States Office of Foreign Disaster Assistance.

Received 20 July 1987; accepted 12 February 1988.

- Karig, D. E. *J. geophys. Res.* **76**, 2542-2561 (1971).
- Moberly, R. *Geol. Soc. Am. Mem.* **132**, 35-55 (1972).
- Taylor, B. & Karner, G. D. *Rev. Geophys. Space Phys.* **21**, 1727-1741 (1983).
- Weissel, J. K. *Phil. Trans. R. Soc. A300*, 223-247 (1981).
- Lawver, L. A., Hawkins, J. W. & Schlater, J. G. *Earth planet. Sci. Lett.* **33**, 27-35 (1976).
- Lawver, L. A. & Hawkins, J. W. *Tectonophysics* **45**, 323-339 (1978).
- Tamaki, K. *Geology* **13**, 475-478 (1985).
- Chase, C. G. *Geol. Soc. Am. Bull.* **82**, 3087-3110 (1971).
- Malahoff, A., Feden, R. H. & Fleming, H. S. *J. geophys. Res.* **87**, 4109-4125 (1982).
- Isacks, B. L., Sykes, L. R. & Oliver, J. *Geol. Soc. Am. Bull.* **80**, 1443-1470 (1969).
- Hamburger, M. W. & Everingham, I. B. R. *Soc. N.Z. Bull.* **24**, 439-453 (1986).
- Hamburger, M. W. & Isacks, B. L. in *Circum-Pacific Council Energy Mineral Resources Earth Science Ser.* (eds Kroenke, L. W. & Eade, J. V.) (Circum-Pacific Council for Energy and Mineral Resources, Houston, Texas, in the press).
- Dewey, J. F. *Am. J. Sci.* **275A**, 260-284 (1975).
- Minster, J. B. & Jordan, T. H. *J. geophys. Res.* **83**, 5331-5354 (1978).
- Schlater, J. G., Hawkins, J. W., Mammerickx, J. & Chase, C. G. *Geol. Soc. Am. Bull.* **83**, 505-517 (1972).
- Hawkins, J. W. in *The Geology of Continental Margins* (ed. Burke, C. A.) 505-520 (Springer, New York, 1974).
- Hawkins, J. W. & Melchior, J. T. *J. geophys. Res.* **90**, 11431-11468 (1985).
- Aggarwal, Y. P., Barazangi, M. & Isacks, B. *J. geophys. Res.* **77**, 6427-6434 (1972).
- Dubois, J., Pascal, G., Barazangi, M., Isacks, B. L. & Oliver, J. *J. geophys. Res.* **78**, 3431-3436 (1973).
- Barazangi, M., Isacks, B. L., Dubois, J. & Pascal, G. *Tectonophysics* **24**, 1-12 (1974).
- Auzende, J. M., Eissen, J. P., Lafoy, Y., Gente, P. & Charlou, J. L. *Tectonophysics* (submitted).
- Kroenke, L. W., Smith, R. & Nemoto, K. in *Circum-Pacific Council Energy Mineral Resources Earth Science Ser.* (eds Kroenke, L. W. & Eade, J. V.) (Circum-Pacific Council for Energy and Mineral Resources, Houston, Texas, in the press).
- Brocher, T. M. & Holmes, R. in *Circum-Pacific Council Energy Mineral Resources Earth Science Ser.* 3 (ed. Brocher, T. M.) 122-127 (Circum-Pacific Council for Energy and Mineral Resources, Houston, Texas, 1985).
- Hawkins, J. W. & Batiza, R. *Eos* **56**, 1078 (1975).
- Halunen, J. thesis, Univ. Hawaii (1979).
- Von Stackelberg, U. et al. *BGR Circular* **2**, 3-14 (1985).
- Weissel, J. K. in *Island Arcs, Deep Sea Trenches, and Back-Arc Basins* (eds Talwani, M. & Pitman, W. C.) 429-436 (American Geophysical Union, Washington, DC, 1977).
- Morton, J. L. & Sleep, N. H. in *Circum-Pacific Council Energy Mineral Resources Earth Science Ser.* 2 (eds Scholl, D. W. & Vallier, T. L.) 441-456 (Circum-Pacific Council for Energy and Mineral Resources, Houston, Texas, 1985).
- Hamburger, M. W. thesis, Cornell Univ. (1986).
- Sykes, L. R., Isacks, B. L. & Oliver, J. *Bull. seism. Soc. Am.* **59**, 1093-1113 (1969).

31. Dziewonski, A. M., Chou, T. A. & Woodhouse, J. H. *J. geophys. Res.* **86**, 2825–2852 (1981).
32. Eguchi, T. *Tectonophysics* **102**, 17–32 (1984).
33. McKenzie, D. P. *Bull. seism. Soc. Am.* **59**, 591–601 (1969).
34. Harding, T. P., Vierbuchen, R. C. & Christie-Blick, N. *Soc. Econ. Paleontol. Mineral. Spec. Publ.* **37**, 51–81 (1985).
35. Kroenke, L. W., Jouannic, C. & Woodward, P. *Geophysical Atlas of the Pacific*, Chart 1 (CCOP/SOPAC, Wellington, N.Z., 1983).
36. Hamburger, M. W., Everingham, I. B., Isacks, B. L. & Barazangi, M. *Geology* **16**, 237–241 (1988).
37. Rodda, P. & Kroenke, L. W. *U.N. ESCAP CCOP/SOPAC Tech. Bull.* **6** (ed. Kroenke, L. W.) 87–109 (1984).
38. Falvey, D. A. *Bull. Aust. Soc. Explor. Geophys.* **9**, 117–123 (1978).
39. Malahoff, A., Hammond, S. R., Naughton, J. J., Keeling, D. L. & Richmond, R. N. *Earth planet. Sci. Lett.* **57**, 398–414 (1982).
40. Crowell, J. C. *Soc. Econ. Paleontol. Mineral. Spec. Publ.* **22**, 190–204 (1974).
41. Mann, P., Hempton, M. R., Bradley, D. C. & Burke, K. *J. Geol.* **91**, 529–554 (1983).
42. Dibblee, T. W. Jr in *Late Mesozoic and Cenozoic Sedimentation and Tectonics in California* 26–38 (San Joaquin Geol. Soc., Bakersfield, CA, 1977).
43. Barker, P. F. & Hill, I. A. *Phil. Trans. R. Soc. A300*, 249–262 (1981).
44. Taylor, B. *Geology* **7**, 171–174 (1979).
45. Hussong, D. M. & Uyeda, S. *Init. Rep. DSDP* **60**, 909–929 (1981).
46. Hussong, D. M. & Sinton, J. B. in *The Tectonic and Geologic Evolution of Southeast Asian Seas and Islands* Vol. 2 (ed. Hayes, D. E.) 217–235 (American Geophysical Union, Washington, DC, 1983).
47. Lallemant, S. & Jolivet, L. *Earth planet. Sci. Lett.* **76**, 375–389 (1986).
48. Biju-Duval, B., Letouzey, J. & Montadert, L. in *Geological and Geophysical Investigations of Continental Margins* (eds Watkins, J. S., Montadert, L. & Dickerson, P. W.) 293–317 (American Association of Petroleum Geologists, Tulsa, Oklahoma, 1979).
49. *Plate Tectonic Map of the Circum-Pacific Region* (American Association of Petroleum Geologists, Tulsa, Oklahoma, 1981).
50. Heezen, B. C. *Geol. Soc. Am. Spec. Pap.* **65**, Plate 1 revised (1968).

Two-tiered regulation of spatially patterned engrailed gene expression during *Drosophila* embryogenesis

Stephen DiNardo*, Elizabeth Sher, Jill Heemskerk-Jongens, Judith A. Kassis* & Patrick H. O'Farrell

Department of Biochemistry and Biophysics, UCSF School of Medicine, San Francisco, California 94143-0448, USA

A regulatory cascade, initiated during the syncytial stage of embryogenesis, culminates in the striped pattern of engrailed gene expression at the cellular blastoderm stage. The early regulatory genes, for example the pair-rule genes, are expressed transiently and as their products decay a distinct regulatory programme involving segment polarity genes takes over. This late programme maintains and perhaps modifies the striped pattern of engrailed expression through interactions that may involve cell communication.

AN early step in the generation of the *Drosophila* body pattern is the determination of the segmental subdivisions of the embryo. This occurs quite rapidly. By the cellular blastoderm stage (2–2.5 h after fertilization) the segmental outlines of the embryo, although not morphologically visible, are represented by the precise expression patterns of the segmentation (or segment polarity) genes, such as *engrailed*^{1,2} and *wingless*³. The subsequent activity of genes such as these directs the morphological events leading to segmentation^{4,5}.

The segmentation genes are expressed in their localized patterns as a consequence of an early regulatory cascade. Three gene classes (maternal genes⁶, and the zygotically acting gap and pair-rule genes⁴) act in sequence, each establishing the more refined spatial expression pattern of the next^{7–9}. This takes place in a syncytial cytoplasm and is thought to rely on the diffusion and interaction of regulatory products encoded by these genes. The subdivision of the embryonic field by this cascade is rapid, occurring during the first thirteen nuclear divisions. After the thirteenth division, the nuclei are cellularized and the cellular blastoderm is formed.

One outcome of this regulatory hierarchy is the establishment of *engrailed* (*en*) expression at the cellular blastoderm stage in 14 single-cell-wide stripes transecting the antero-posterior axis of the embryo^{1,2,10,22}. Previous work has shown that the activity of separate sets of pair-rule genes controls establishment of *en* expression in even- and in odd-numbered stripes^{11,12}. The data presented here suggests further that the separate control of even- and odd-numbered stripes is reflected in the organization of the cis-regulatory sequences at *en*.

During the post-cellular blastoderm phase of development the expression pattern of *en* is maintained and modified^{1,10}. It is unlikely that the pair-rule genes are involved in the post-

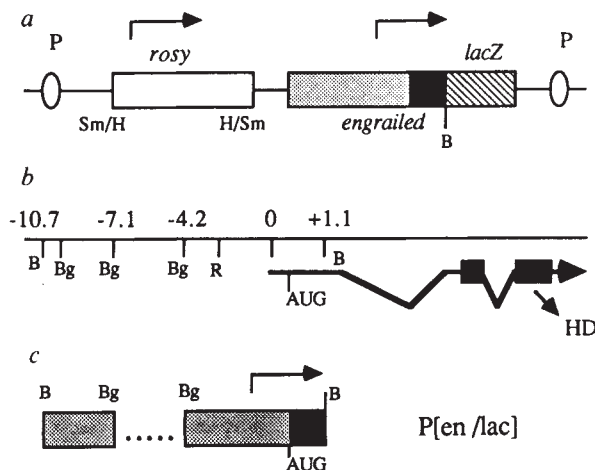


Fig. 1 Construction of an *en-lacZ* fusion gene (not to scale). *a*, schematic diagram of *P(en/lac)* indicating sites for P-mediated integration (P); the *rosy* gene (open box); upstream and non-coding (stippled) and coding (filled) regions from *engrailed*; and *lacZ* coding sequences (hatched). An SV40 polyadenylation signal is at the end of *lacZ*. *b*, Map of *en* genomic region showing particular restriction site landmarks, the major transcription unit with start site (0) and presumed translation start site (AUG) and the portions of exons two and three (solid boxes) encoding the homoeodomain (HD)^{19,20,40}. *c*, Specific genomic *en* regions used in constructing *P(en/lac)* (aligned with *b*). Dots represent the deletion of sequences between coordinates -4.2 and -7.1. Restriction sites: B, *Bam*HI; Bg, *Bgl*II; H, *Hind*III; R, *Eco*RI; Sm, *Sma*I; symbol (/), joined by blunt-end ligation. Arrows indicate the direction of transcription.

Methods. Transgenic lines were established¹⁸ using 'wings-clipped' helper, and an *Sgs-4^{Be-1}*, *cn*; *ry⁵⁰⁶* host. Lines were established from three independent *G₀* adults; the insertion mapped to chromosome II in one line and to chromosome III in the other two.

*Present addresses: The Rockefeller University, 1230 York Ave., New York, New York 10021, USA (S. DiN.); Department of Health and Human Services, PHS FDA, Bethesda, Maryland 20892, USA (J.A.K.).

blastoderm regulation of *en*, since the pair-rule genes characterized to date are only expressed transiently in their pair-rule blastoderm patterns^{13-16,31}. Here we address the mechanism by which the *en* gene is regulated in later embryonic development.

We find that the striped expression pattern of *en* during post-blastoderm development is controlled by a cis-regulatory programme distinct from that controlling the establishment of expression at the cellular blastoderm stage. Control of this late-acting programme is exercised through the activity of various segment polarity genes. The activity of at least one of the segment polarity genes (*wingless*) is not cell autonomous^{17,44}. This leads us to a model for the control of post-blastoderm *en* gene expression involving cell interactions. We suggest that the late *en* regulatory programme allows for some plasticity in developmental programming.

An *en-lacZ* fusion gene

As part of an ongoing dissection of the *en* regulatory region we have constructed transgenic fly lines expressing β -galactosidase under the control of portions of the *en* regulatory region. Our dissection of the cis-regulatory sequences at *en* is in its initial stages, but the expression pattern of one transgenic construct is particularly instructive. It accurately expresses discrete portions of the *en* pattern and represents one line of evidence that there are distinct early and late regulatory programmes that control the expression of *en* in stripes.

Figure 1 shows the DNA construct, P(*en/lac*), in which expression of a β -galactosidase (β -gal) - engrailed fusion protein is directed by the *en* promoter and flanking sequences. The fusion gene is in a vector allowing the establishment of stably transformed 'transgenic' flies through P-element mediated integration into the germ line¹⁸. In addition to the promoter and 5' flanking sequences, the construct includes 1099 base pairs (bp) of transcribed sequences from the *en* gene¹⁹, encompassing ~211 nucleotides of untranslated leader and the N-terminal 296 amino acids²⁰ spliced in frame to the *lacZ* gene encoding β -galactosidase (β -gal) activity. Therefore, the expression of this construct could reflect control both at the transcriptional and post-transcriptional levels. The signal (β -gal activity or antigen) produced by P(*en/lac*) is unstable, thereby allowing us to detect transitions that occur in *en* expression patterns. The instability could be due to the inclusion of *en* leader and coding sequences in the fusion, as a relatively stable β -gal signal is produced by a *ftz* promoter-*lacZ* fusion, which does not contain coding sequences from the *ftz* gene²¹.

Early expression

Three independent P(*en/lac*) lines have been analysed, and all exhibit a similar pattern of ectodermal expression. Seven stripes of expression from P(*en/lac*) are apparent along the extending germband (Fig. 2a), a stage at which *en* is expressed in fourteen stripes. The seven stripes expressing P(*en/lac*) precisely align with the even-numbered *en* stripes (Fig. 2b, c). Within these stripes P(*en/lac*) is faithfully expressed since all cells expressing *en* are also expressing β -gal. On occasion, an extra β -gal-expressing cell is observed ventrally at the trailing edge of even-numbered *en* stripes (Fig. 2b, c). The significance of this occasional non-correspondence is unclear.

The establishment of P(*en/lac*) expression at even-numbered *en* stripes is first evident at the same stage as endogenous *en* expression—the cellular blastoderm/early gastrula (Fig. 2d, e). As endogenous *en* expression is established, the odd-numbered stripes initially lag behind the development of even-numbered stripes^{10,22}. However, *en* is equivalently expressed in the odd- and even-numbered stripes by the time the germband is elongating (compare panels d and b, Fig. 2). In contrast, during germband elongation or even after the germband fully extends, no β -gal expression (enzyme or antigenic activity) is detected in positions corresponding to the odd-numbered *en* stripes (Fig. 2a).

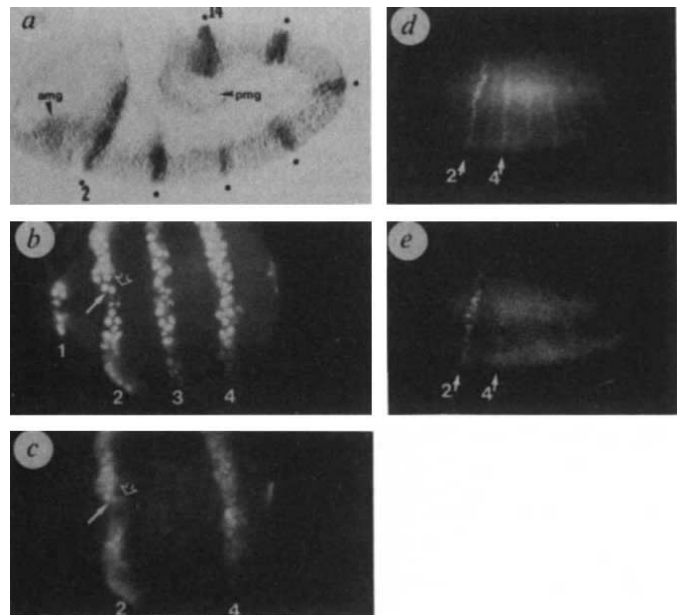


Fig. 2 Antibody-staining of P(*en/lac*) transgenic embryos show that regulatory sequences controlling establishment of even-numbered *en* stripes are functionally distinct from those controlling establishment of odd-numbered stripes. Anterior left and ventral down in all panels, unless indicated otherwise. **a**, P(*en/lac*) transformant stained with rabbit anti- β -galactosidase antibody to visualize the expression from P(*en/lac*), 3.5–4 h after egg laying (AEL). Seven ectodermal stripes are evident (indicated by dots), corresponding to *en* stripes 2, 4, 6, 8, 10, 12 and 14, which mark the posterior portions of the maxillary, first and third thoracic, and second, fourth, sixth and eighth abdominal primordia. Variable expression seen in the anterior (amg) and posterior (pmg) midgut primordia is not part of the normal *en* expression program. This expression may be due to sequences from *en*, or may be spurious, since expression from similar vectors in mesodermal and gut primordia has been reported⁴¹. **b**, **c**: Magnified (ventral) view of a doubly-labelled embryo (mouse anti-*en* and rabbit anti- β -gal) reveals expression of endogenous *en* (**b**) and of P(*en/lac*) (**c**), 4.5 h AEL. The stripes of *en* expression are 1(Mn), 2(Max), 3(La) and 4(T1). Expression of β -gal is coincident with even-numbered *en* stripes 2(Max) and 4(T1); arrows indicate the same cells in **b** and **c**. There is no detectable β -gal expression at positions corresponding to odd-numbered *en* stripes. In even-numbered stripes there is virtually one-to-one correspondence between β -gal and *en* expressing cells. The occasional cell containing low levels of β -gal antigen but no detectable *en* is indicated (open arrow). The β -gal signal is not as discrete as the *en* signal since the *en*- β -gal fusion protein is not restricted to the nucleus. **d**, **e**: Embryos were doubly-labelled for *en* expression (**d**) and β -gal expression (**e**) at the onset of gastrulation, 3 h AEL. Expression from P(*en/lac*) is induced at about the same developmental stage as *en*, although there may be a slight lag in induction or accumulation. Arrows point to co-expression in 2(Max), and in 4(T1), where β -gal is just detectable.

Methods. Preparation of embryos for immunocytochemistry was as in refs 10, 12. Rabbit anti- β -gal was from Cappel, rabbit anti-*en* (ref. 10); mouse monoclonal anti-*en* (and inv), a gift of K. Coleman, C. Goodman and T. Kornberg. Final magnification was $\times 130$, except for **b** and **c** which were $\times 520$.

To test whether the establishment of P(*en/lac*) expression is under the control of the same regulatory factors that control the establishment of even-numbered *en* stripes, we have analysed the expression of P(*en/lac*) in particular pair-rule mutants. Two pair-rule mutants known to affect the expression of even-numbered *en* stripes, *fushi tarazu*^{11,12} and *odd-skipped*¹², have identical effects on the establishment of P(*en/lac*) expression and on endogenous *en* expression (data not shown).

We conclude that P(*en/lac*) expression is regulated faithfully at positions corresponding to even-numbered *en* stripes, and

Fig. 3 P(en/lac) uncovers a distinct late program of control over *en*, which is regulated by segment polarity genes. *a*, staining for β -gal, 5 h AEL, anterior left and ventral down ($\times 130$). The seven β -gal stripes (dots) are beginning to fade. Note the spotty appearance laterally (open arrows). As development proceeds this signal fades ventrally. In addition, note that some cells at dorso-lateral positions are beginning to accumulate β -gal signal (solid arrows). The signal in the dorsal ectoderm will be heavily induced during subsequent stages. Though restricted to dorsal regions, the position of this signal corresponds accurately to even- and odd-numbered *en* stripes. *b*, Staining for β -gal at the onset of germband retraction, 7–8 h AEL, ventro-lateral view ($\times 130$). Expression of P(en/lac) in ventral areas is low, but expression at dorso-lateral positions is quite high. Expression in both ventral and dorsal positions occurs in the posterior portions of 11 segment primordia: Labial, T1–T3, and A1–A7. Significant expression of P(en/lac) in Mandibular and Maxillary primordia and the terminalia during this late programme is usually not observed. Expression in dorsal-most A8 is sometimes observed. Bracketed area is magnified (from a different embryo) in *c* and *d*. VM, ventral midline. *c*, *d*: Magnified view ($\times 520$) of a doubly-labelled embryo at the turn of the germband (A2, A3 and A4 region) showing one to one correspondence between *en* (*c*) and β -gal (*d*) expressing cells. *e*, *f*: Same as *c*, *d* but in *ptc* mutant background, with arrow showing that P(en/lac) is ectopically induced (*f*) as is the endogenous *en* gene (*e*).

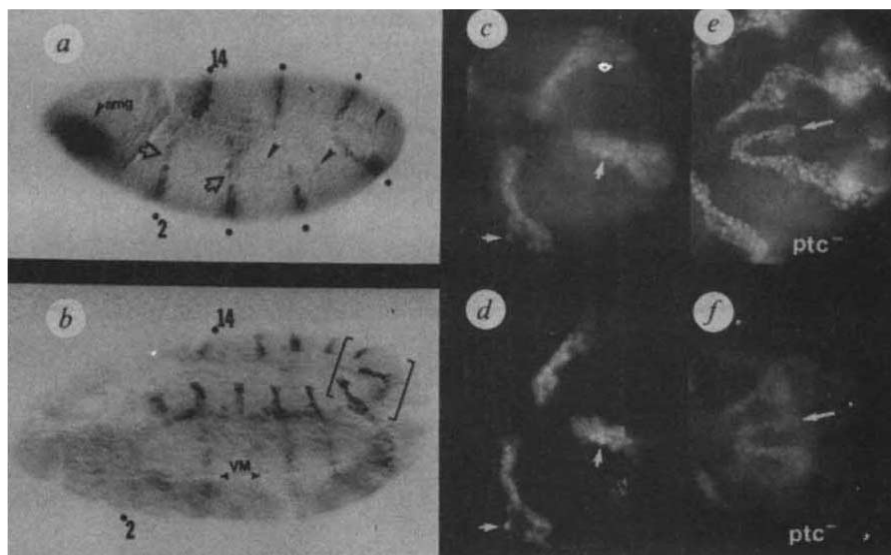
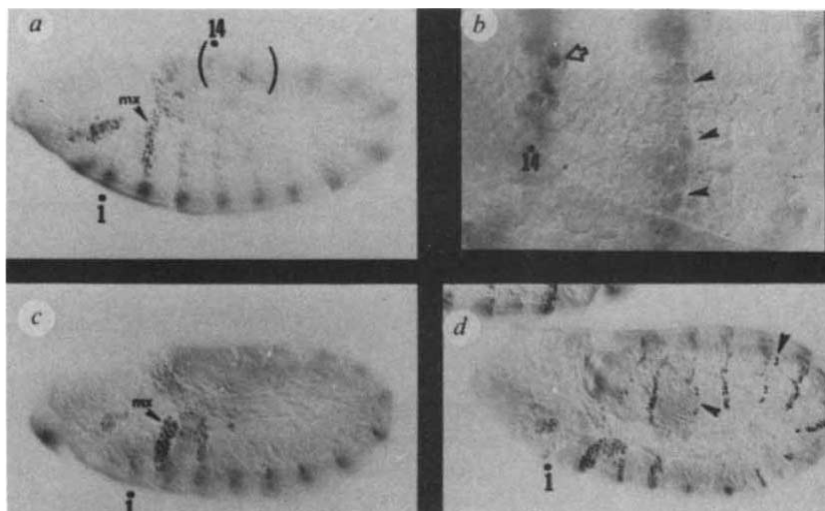


Fig. 4 Premature decay of *en* expression in *wg* and *hh* mutants. Embryos, anterior left and ventral down, were stained with anti-*en*-antibody. *a*, Mutant *wg*^{cx4}: 4–5 h AEL ($\times 130$). Most *en* stripes are faint, but still visible above background. The Max and La stripes are affected least. Similar results have been obtained in *wg*^{L5}, *wg*^{IG22}, and also in *wg*^{IL114ts} when aged at non-permissive temperatures (25–28 °C). In wild-type embryos *en* antigen can be visualized in cells until the onset of cuticular differentiation, 12–14 h AEL¹⁰. *b*, Mutant *wg*^{cx4}: magnified ($\times 520$) view (in a different embryo) of bracketed region shown in (*a*) as signal from *en* expression decays. Particular *en* stripes lose signal at varying times; thus, these two adjacent *en* stripes clearly show that the decay in *en* expression occurs while cells are still part of the ectodermal cell sheet (arrowhead, decaying signal). *c*, Mutant *wg*^{cx4}: 6.5 h AEL ($\times 130$). The remaining *en* signal is lost from ectodermal cells, except in Max and La regions. The blurred ventral signal is due to *en* expression in cells (out of this focal plane) of the developing central nervous system. *d*, *hh*^{6N}: 6 h AEL ($\times 130$). Note the spotty and thin appearance of *en* stripes. Similar results are found for *arm*^{xk22}, *fu*¹¹³, *fu*^{113/fu}⁹⁴, *hh*^{6N/hh}^{9K}, and in *wg*^{IL114ts} (raised at 16–18 °C). The segment polarity mutants *gooseberry* and *cubitus interruptus*^D have not been studied.



that this regulation involves the same trans-acting activities as the endogenous *en* gene. Therefore, the regulatory region present in P(en/lac) includes cis-acting sequences that direct the establishment of even-numbered *en* stripes. Presumably, the sequences required for expression in odd-numbered stripes are not present, or are functionally disrupted in this construct.

Late expression

As development continues, there is a striking transition in the expression of P(en/lac). The changes in expression provide evidence that the expression of *en* stripes is separable into two temporally and mechanistically distinct programmes.

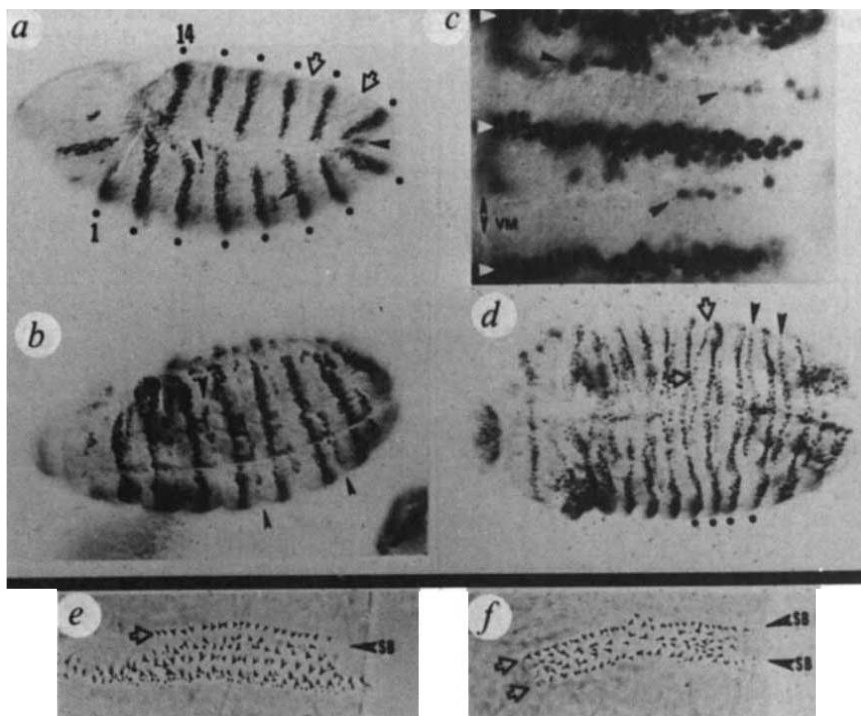
After the germband is fully extended, the expression of P(en/lac) in stripes corresponding to even-numbered *en* stripes begins to fade. (Note the spotty appearance of the stripes in Fig. 3*a* where the embryo is about 1.5 h older than the stage exhibited in Fig. 2*a*.) At this late stage, individual cells located dorso-laterally at positions corresponding to odd-numbered *en* stripes begin to accumulate β -gal antigen (Fig. 3*a*). Over the

next hour or two, β -gal staining is lost ventrally from even-numbered stripes, but intense expression is induced dorso-laterally at positions corresponding to both even- and odd-numbered *en* stripes (in the dorsal ridge, T1 through A7, and to some extent A8; Fig. 3*b*). The eleven dorso-lateral stripes expressed from P(en/lac) at this late developmental stage correspond cell-for-cell with *en* expressing cells (Fig. 3*c*, *d*).

As the germband retracts the P(en/lac) signal appears to extend ventrally, although it usually does not completely fill the *en*-expressing region. Often, the level of P(en/lac) expression on the ventral surface is higher in cells at the posterior edge of the *en* stripes, and lower in other cells of the *en* stripe (data not shown). In contrast, dorso-lateral expression of P(en/lac) remains uniformly coextensive with *en*-expressing cells. We conclude that the P(en/lac) construct contains at least some of the regulatory sequences necessary for position specific expression in ventro-lateral regions, whereas it contains all sequences necessary for dorso-lateral expression. The discrimination between dorso-lateral and ventro-lateral regions in the

Fig. 5 Reprogramming of cells to express *en* in *patched* mutants. Mutant embryos *ptc*^{IN108} were stained with anti-*en*-antibody. *a*, Germ-band extended embryo (4.5 h AEL), anterior left and ventral down ($\times 130$). The normal 14 *en* stripes, (established at the cellular blastoderm stage) are now two to three cells wide and are indicated by dots. The patches and stripe of *en*-expressing cells anterior to stripe 1 resemble wild-type. Cells newly accumulating *en* antigen are indicated (arrowheads). The stage at which ectopic expression is first detectable cannot be due to leakiness of the *ptc* allele because the same delay is observed in *ptc*^{IN108} homozygotes as in the deficiency/*ptc*^{IN108} heteroallelic combination. A furrow in the ectodermal cell sheet is becoming visible between each pair of *en* stripes at the approximate position of ectopic induction (open arrows). The asterisk marks a cell group that induces *en* at this stage in wild-type also. *b*, Ventral view, 6 h AEL ($\times 130$). Eight *en* stripes are visible. More cells now express *en* between each pair of normal stripes, giving the appearance of an eccentrically positioned, incomplete stripe (arrowheads). *c*, Magnified ($\times 520$) ventro-lateral view (6 h AEL) showing three normal *en* stripes (white arrowheads) and cells that have induced *en*, located near and often within the deep furrows. *d*, Germband retracted (9 h AEL) anterior left ventral view ($\times 130$). A few normal stripes (dots) and the intervening ectopic stripes are indicated (arrowheads). A given ectopic stripe has a wavy appearance, fusing along part of its length alternately with a posteriorly located *en* stripe, and then ventro-laterally with an anteriorly located stripe (open arrows). This fusion may be the result of shifts in cell position during germband retraction. Such fusions and reorganization of *en* stripes are frequently seen at late developmental stages in many of the segmentation mutants. *e*, Wild-type fourth abdominal denticle belt with border (SB) between third (upper) and fourth (lower) segment indicated. *f*, Fourth abdominal denticle belt in mutant *ptc*^{IN108}/Df(2R)*ptc* showing duplication of segment border (SB), and anterior row of denticles (open arrows). In *ptc*^{6P43}, *en*^{TM99} double mutants (raised at 18–20 °C) both normal and duplicate anterior row denticles do not appear, showing dependence on *en* function.

Methods. Stocks were obtained from C. Nusslein-Volhard and E. Wieschaus, from N. Baker (*wg*^{ex4}), and J. Hooper and M. Scott (Df(2R)*ptc*44C-E). Crosses were performed in standard fashion, at 25 °C on molasses-cornmeal-agar. Embryos were prepared for immunocytochemical detection of *en* protein by affinity-purified rabbit anti-*en* antibody^{10,12}. Cuticles were prepared as in ref. 42.



embryo is not unique to this construct. Recent data have shown that there is also a sharp ventral/dorsal distinction in neurogenic potential and in the mitotic division schedule²³ (V. Foe, personal communication).

At this later developmental stage, the expression of P(*en/lac*) in positions corresponding to odd-numbered *en* stripes is in striking contrast to the absence of detectable β -gal expression in these positions at early stages. Therefore, P(*en/lac*) responds later in development to regulatory signals that control expression in both even- and odd-numbered positions. These data show that two programmes of striped *en* expression are distinguishable by character and time of onset. The first programme results in the initiation of stripes at the cellular blastoderm, and involves the separate regulation of even- and odd-numbered stripes by pair-rule gene action. The second programme becomes evident after the germband has elongated, as initial expression fades and *de novo* expression becomes visible at both even- and odd-numbered positions. Since late P(*en/lac*) expression is not present in all cells that express the endogenous *en* gene (for example, ventral P(*en/lac*) stripes do not fill the *en* domain), it is unlikely that P(*en/lac*) expression is being regulated simply by the presence of endogenous *en* protein. Such a positive autoregulatory loop may be one of the components of late regulation, but *en* must also be responding to other signals. We next address the nature of these other inputs by investigating the effects that particular segmentation genes have on the pattern of expression of the endogenous *en* gene.

Regulation of late programme

It is unlikely that the pair-rule genes directly regulate the late programme of *en* expression, since products encoded by these genes are no longer detectable at this stage^{13–16,24,25,31}. Analysing

the *en* expression pattern in particular segment polarity mutants provides evidence that these genes are involved in regulating the second programme.

In *wg* null mutants *en* expression is established normally at the cellular blastoderm in fourteen roughly single-cell-wide stripes. This pattern remains wild-type through gastrulation and germband elongation, *en* stripes becoming 2–3 cells wide at this stage as a result of both cell movements and cell division. However, after the germband is fully extended, ectodermal *en* expression decays prematurely. Two developmental stages are shown in Fig. 4*a, c*, revealing the progress of the decay (summary, Fig. 6). The only ectodermal expression that remains is in the maxillary (Mx) stripe, a portion of the labial (La) stripe, and in a cluster of dorsally located cells in developing T1 (Fig. 4*c*). The effect of strong, but probably not null, alleles of *hh*, *arm* and *fu* are similar to the effects of the *wg* mutations, except that the *en* stripes do not disappear fully. Rather, a variable but small number of cells retain *en* expression. An example is shown for *hh*^{6N} (Fig. 4*d*) (note the spotty and 'thin' appearance of *en* stripes). Weak alleles of *wg*, such as *wg*^{IL114ts} grown at semi-permissive temperature, similarly leave some *en*-expressing cells unaffected (data not shown). This raises the possibility that the true null condition for *hh*, *fu* and *arm* (including removal of any maternal contribution), would result in the total loss of *en* expression as in *wg* nulls.

Some cell death occurs in these mutants during germband retraction^{26,27}. Thus, it is possible that cell death leads to the observed loss of *en* signal. However, in *wg* mutants, there is no evidence of cell death at the stage when cells are in the process of losing *en* antigen (Fig. 4*b*). This suggests that the loss of *en* expression is an early effect in *wg* mutants and not secondary to cell death.

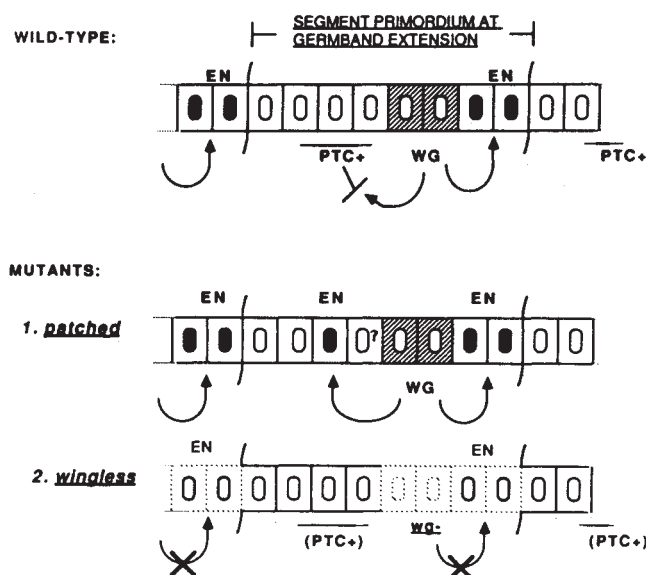


Fig. 6 Interactions regulating *en* during post-blastoderm development. Wild-type: The segmental primordium at germband extension stage (anterior to left) is represented as eight cells wide—the roughly four-cell wide primordia having expanded due to a post-cellular blastoderm cell division and some interdigitation of cells. Several domains within each primordium appear to exist, each 'marked' by the expression of, or by the requirement for the activity of, particular genes. Cells expressing *en* and *wg* account for two of the domains: the *en* domain comprises the last two cells of the primordium⁴⁻⁶ (blackened nuclei); the anteriorly adjacent two cells constitute the *wg* domain³ (hatching). The analysis of *ptc* mutations suggests that the remaining cells of the primordium are divided into at least two domains: those that ectopically induce *en* in the absence of *ptc* activity, and those that do not. The placement of the *ptc*-requiring domain is based on the position of ectopic *en* expression in *ptc* mutants. It has been proposed that the activity of another gene, *naked*, is required for the anterior-most cells of the primordium³⁵. Each of the domains is presumably established at the cellular blastoderm stage by the action of pair-rule segmentation genes, as has been shown for *en*^{11,12} and *wg* domains⁴³. The establishment of *en* expression constitutes the early regulatory programme of *en*. The second regulatory programme acts during post-blastoderm development as *wg*-expressing cells induce (or maintain) *en* expression in posteriorly located cells (arrow). Data presented here do not rule out other interactions. Mutants: (1) In the absence of *ptc* activity, *en* is induced in cells anterior to the *wg* domain. The induction requires *wg* (arrow to anterior), as no induction is detected in the absence of *wg* activity. Thus, in wild-type, *wg* induces *en* posterior to the *wg* domain, but is inhibited, directly or indirectly, from doing so anteriorly by *ptc* activity (blocking arrow shown in 'wild-type' panel). A cell (marked '?') lies between the *wg* domain and the cell ectopically inducing *en* (see text). (2) the *wg* mutation leads to the premature loss of *en* expression (stippled nuclei), as a result of the loss of a morphogenetic signal from *wg* expressing cells, perhaps the *wg* product itself. The *hh*, *fu* and/or *arm* products may play some role in the transduction of this signal. Cuticle pattern normally produced by *en* and *wg*-dependent cell types is missing^{3,4} either due to cell death or transformation of fate (dashed cell outlines). It is not known if the *wg* mutation affects cells located anterior to the *wg* domain.

The results indicate that the segment polarity loci *wg*, *fu*, *hh* and *arm* act in some manner to reinforce or maintain *en* expression once it has been established. The segment polarity mutant *patched* (*ptc*)⁴ has qualitatively different effects on *en* expression, providing further evidence for the existence of a distinct regulatory programme controlling *en* expression during post-blastoderm development.

In *patched* mutants, *en* expression is established normally. However, at full germband extension in *ptc* mutants, the program

of *en* expression is altered as *en* is induced in cells lying a few cell diameters posterior to each normal *en* stripe (Fig. 5a). The cells that are inducing *en* lie near and/or within an abnormal furrow that develops in the ectodermal cell sheet (Fig. 5a, b, c). As development proceeds more cells express *en* (Fig. 5b). By the onset of germband retraction the ectopic expression forms an incomplete transverse stripe positioned eccentrically between each pair of normal *en* stripes (Fig. 5d; summary, Fig. 6).

The *ptc* mutant phenotype suggests that the novel *en* stripes have a functional consequence on development. The *ptc* mutants develop duplicated cuticular structures within each segment⁴. Among these duplicated structures are an anterior row of denticle hairs and a segment border, which are characteristic of *en*-expressing cells²⁸ (S. D., unpublished data; Fig. 5e, f). In a *ptc*, *en* double mutant, whereas other duplicated structures are still produced, those dependent on *en* activity are not produced. We conclude that the ectopic induction of *en* has a consequence on the final body pattern. Therefore, in *ptc* mutants particular cells that are initially specified at the cellular blastoderm stage to be negative for *en* expression are reprogrammed later in development.

wg and ectopic induction

The absence of the *ptc* product leads to a late programme of *en* expression in novel positions within a developing segment. Although pair-rule gene activity is unlikely to control late *en* expression in wild-type embryos, ectopic induction might in principle involve the same regulatory products that establish early *en* expression, such as *ftz* and *eve*, if their expression patterns are correspondingly altered in a *ptc* mutant. We consider this unlikely since we and others find no changes or increase in the lifetime of the *ftz*⁸ and *eve*²⁹ patterns, nor evidence for renewed expression just before, or concomitant with, the induction of ectopic stripes. Furthermore, we can experimentally induce late expression of *ftz* in transgenic flies that express *ftz* from the inducible heat shock promoter³⁰ and such late *ftz* expression has no consequence on the late *en* pattern. The observations suggest that these pair-rule gene products are neither present nor competent to induce ectopic *en* expression at the time that ectopic *en* stripes appear in *ptc* mutants.

In fact, we find that ectopic induction of *en* in *ptc* mutants is dependent on *wg* activity. Specifically, in *wg*, *ptc* double mutants, *en* expression is established normally, but decays prematurely, as it does in *wg* single mutants. There is no subsequent ectopic induction as seen in *ptc* single mutants (data not shown).

The ectopic induction of *en* in positions where there was no previously detectable *en* accumulation raises the possibility that *wg* activity has the capacity to induce *en* expression *de novo*. Therefore, what we have been referring to as 'maintenance' of *en* expression by *wg*, and other segment polarity genes, may be a re-induction of *en* expression. This observation complements the conclusion drawn from the analysis of P(*en/lac*) expression, which suggested that the *en* program has a distinct, late regulatory component that is independent of the establishment of early expression.

Late control of *en-lacZ*

Interestingly, the developmental stage at which segment polarity gene mutations affect *en* expression coincides with the transition between early and late *en* regulatory programmes as determined from the analysis of P(*en/lac*) expression (see above). The temporal link implies that the late programme of induction from P(*en/lac*) is due to the action of segment polarity genes. This has been tested by examining P(*en/lac*) expression in *wg* and *ptc* mutant embryos. In *wg* mutants, early P(*en/lac*) expression is identical to that in wild-type embryos, but the late programme does not initiate (data not shown). In *ptc* mutants, the early programme is also normal, but when the late programme is induced, P(*en/lac*) is expressed ectopically, as is the endogenous *en* gene (Fig. 3e, f). In summary, the effects of *wg* or *ptc* muta-

tions on expression from *P(en/lac)* correlate with the effect of these mutations on endogenous *en* expression.

Cell communication

The data show that there are two programmes regulating *en* gene expression in stripes during embryogenesis. The first regulatory programme involves the independent control of even- and odd-numbered *en* stripes by pair-rule gene action^{11,12,15}, and may involve direct transcriptional regulation by some of the pair-rule products.

The second regulatory programme operates after cellularization, and involves segment polarity gene action. Recent molecular analysis of *wg*, coupled with the data presented here, suggests that cell interactions are involved in the late control of *en* striped expression.

The *wg* gene, like *en*, is expressed in stripes, but *wg* transcripts accumulate mainly in cells that are anteriorly adjacent to those that express *en*, and probably not in the cells that express *en* (ref. 3, Fig. 6). Similarities noted between the predicted *Drosophila* *wg* protein and the product of the mammalian oncogene *int-1* have led to the suggestion that the *wg* product acts as a signal in cell communication³². This suggestion is consistent with genetic analysis showing that *wg* is not cell-autonomous^{17,33,44}. Our results identify one set of target cells for *wg* activity—the neighbouring *en*-expressing cells (Fig. 6)—and we suggest that *en* gene expression may be a target of a signal that is initiated at the cell surface by the *wg* protein. Since *hh*, *arm* and *fu* have phenotypes similar to *wg* with respect to their effects on *en* expression, it is tempting to speculate that some of these genes may encode other components involved in the transduction of the *wg* signal. The existence of genes in the mouse that are homologous to *en*³⁴ and *wg*³² suggests that the circuitry proposed here for the control of *en* by *wg* may be phylogenetically conserved.

In formal terms *ptc* activity is required to keep *en* expression repressed in particular cell rows within each developing segment. Interestingly, the derepression of *en* in a *ptc* mutant requires a signal from *wg*-expressing cells. Consequently, one function of *ptc* may be to prohibit, directly or indirectly, *wg* from inducing *en* ectopically (Fig. 6). Thus, *ptc* activity may impart a polarity on patterning within segments, since the induction of *en* is restricted to cells located posterior to the *wg* domain (Fig. 6).

Ectopic induction of *en* in *ptc* mutants occurs in cells which do not abut the normal *wg* domain. Either *wg* can act over several cell diameters, or there are other genetic interactions missing from the scheme. Indeed, Martinez-Arias and co-workers have found that the *wg* expression domain is expanded in *ptc* mutants and that this change precedes the induction of *en*³⁵. Thus, the defects in a *ptc* mutant may lead to a cascade

of compensatory changes, at least in part involving *wg*, that lead to the induction of *en*.

Separate regulatory programmes

A teleological rationale for the existence of two discrete programmes for striped *en* expression can be offered. The observation that cells are respecified in *ptc* mutants suggests that there is still considerable plasticity in developmental programming at the cellular blastoderm. The later programme, involving the action of segment polarity genes, may represent an editing function that corrects subtle mistakes made in the establishment of *en* expression. In this scenario, the late programme is an important accessory programme that improves the fidelity of the system. In fact, although the establishment of *en* expression at blastoderm is remarkably precise—>95% of the cells in even-numbered *en* stripes map precisely to the anterior-most cells of *ftz* stripes^{36,37}—the remaining 2–5% of cells may represent true 'mistakes'³⁷, where *en* expression is initiated in cells internal to the anterior edge of the *ftz* stripe. Perhaps these mistakes are corrected during the post-blastoderm phase of development. In addition, this editing function may be what governs the classical 'regulative response', in which adjustments are made in body pattern in response to physical perturbations^{38,39}.

A second possibility, not incompatible with the first, is that the existence of a separate early programme is an evolutionary adaptation for the very rapid embryonic development observed in *Drosophila*. The early programme, which establishes striped *en* expression at cellular blastoderm, represents the successful conclusion of a regulatory cascade that subdivides the embryonic field very rapidly⁹ and operates largely in a syncytium. Such a rapid and syncytial developmental plan is rather peculiar. Perhaps the late regulatory programme, involving cell interactions, is sufficient to achieve the gradual and sequential subdivision of the germband seen in most organisms. In this scenario, the late programme would be phylogenetically conserved, and the rapid cascade of regulatory interactions that culminates in the early program would be an accessory programme used by *Drosophila* and its relatives to rapidly establish spatial subdivisions.

We are indebted to Gus del Puerto for help in establishing transgenic lines and immunocytochemistry; P. Ingham, A. Martinez-Arias and N. Baker for communicating unpublished data and members of the laboratory for comments. We thank S. Carroll and M. Frasch for antibodies against *ftz* and *eve*, respectively; J. Hooper and M. Scott for *Df(2R) ptc* stock; and V. Finkel. Research was supported by NSF and NIH (P. O'F.), a Lucille P. Markey Scholar Award (S.D.), Damon Runyon-Walter Winchell (E.S.), NIH training grant (J.H.-J.), and the Giannini Foundation (J.A.K.).

Received 9 December 1987; accepted 1 March 1988.

- Kornberg, T. *et al.* *Cell* **40**, 45–53 (1985).
- Fjose, A., McGinnis, W. J. & Gehring, W. J. *Nature* **313**, 284–289 (1985).
- Baker, N. *EMBO J.* **6**, 1765–1773 (1987).
- Nüsslein-Volhard, C. & Wieschaus, E. *Nature* **287**, 795–801 (1980).
- Kornberg, T. *Proc. natn. Acad. Sci. U.S.A.* **78**, 1095–1099 (1981).
- Nüsslein-Volhard, C., Frohnhöfer, H. G. & Lehmann, R. *Science* **238**, 1675–1681 (1987).
- Gaul, U. & Jäckle, H. *Cell* **51**, 549–555 (1987).
- Carroll, S. B. & Scott, M. P. *Cell* **45**, 113–126 (1986).
- Scott, M. P. & O'Farrell, P. H. *Rev. Cell Biol.* **2**, 49–80 (1986).
- DiNardo, S. *et al.* *Cell* **43**, 59–69 (1985).
- Howard, K. & Ingham, P. *Cell* **44**, 949–957 (1986).
- DiNardo, S. & O'Farrell, P. H. *Genes & Dev.* **1**, 1212–1225 (1987).
- Hafen, E., Kuroiwa, A. & Gehring, W. J. *Cell* **37**, 833–841 (1984).
- Macdonald, P. M., Ingham, P. & Struhl, G. *Cell* **47**, 721–734 (1986).
- Harding, K. *et al.* *Science* **233**, 953–959 (1986).
- Kilcherr, F., Baumgartner, S., Bopp, D., Frei, E. & Noll, M. *Nature* **321**, 493–499 (1986).
- Morata, G. & Lawrence, P. A. *Dev. Biol.* **56**, 227–240 (1977).
- Spradling, A. C. & Rubin, G. M. *Science* **218**, 341–347 (1982).
- Drees, B. *et al.* *EMBO J.* **6**, 2803–2809 (1987).
- Poole, S. J. *et al.* *Cell* **40**, 37–43 (1985).
- Hiroimi, Y., Kuroiwa, A. & Gehring, W. J. *Cell* **43**, 603–613 (1985).

- Weir, M. & Kornberg, T. *Nature* **318**, 433–439 (1985).
- Campos-Ortega, J. A. & Hartenstein, V. *The Embryonic Development of Drosophila melanogaster* (Springer, Berlin, 1985).
- Carroll, S. B. & Scott, M. P. *Cell* **43**, 47–57 (1985).
- Frasch, M. *et al.* *EMBO J.* **6**, 749–759 (1987).
- Martinez-Arias, A. *J. Embryol. exp. Morph.* **87**, 99–114 (1985).
- Perrimon, N. & Mahowald, A. P. *Dev. Biol.* **119**, 587–600 (1987).
- Gergen, J. P., Coulter, D. & Wieschaus, E. in *Gametogenesis and the Early Embryo* (ed. J. Gall 195–220) (New York, 1986).
- Frasch, M. & Levine, M. *Genes & Dev.* **1**, 981–995 (1987).
- Struhl, G. *Nature* **318**, 677–680 (1985).
- Ingham, P. W., Howard, K. R. & Ish-Horowitz, D. *Nature* **318**, 439–445 (1985).
- Rijsewijk, F. *et al.* *Cell* **50**, 649–657 (1987).
- Wieschaus, E. & Riggleman, R. *Cell* **49**, 177–184 (1987).
- Joyner, A. L. & Martin, G. R. *Genes & Dev.* **1**, 29–38 (1987).
- Martinez-Arias, A., Baker, N. & Ingham, P. W. *Development* (submitted).
- Lawrence, P. A., Johnston, P., Macdonald, P. & Struhl, G. *Nature* **328**, 440–442 (1987).
- Carroll, S. B. *et al.* *Genes & Dev.* **2** (in the press).
- French, V., Bryant, P. J. & Bryant, S. *Science* **193**, 969–981 (1976).
- Wright, D. A. & Lawrence, P. A. *Dev. Biol.* **85**, 317–327 (1981).
- Kuner, J. M. *et al.* *Cell* **42**, 309–316 (1985).
- Hiroimi, Y. & Gehring, W. *Cell* **50**, 963–974 (1987).
- Van der Meer, S. *Drosoph. Inf. Serv.* **52**, 160 (1977).
- Ingham, P. W., Baker, N. & Martinez-Arias, A. *Nature* **331**, 73–75 (1988).

Gamma-rays from superconducting cosmic strings

Alexander Vilenkin

Physics Department, Tufts University, Medford,
Massachusetts 02155, USA

Superconducting cosmic strings¹, predicted in some elementary particle theories, can have dramatic cosmological consequences²⁻⁶. Ostriker, Thompson and Witten (OTW) have shown³ that the low-frequency electromagnetic radiation from superconducting closed loops can sweep away the surrounding plasma, and have identified the resulting bubbles of radiation with the voids observed in the large-scale distribution of galaxies. Here I discuss some observational consequences of the OTW model. It will be shown that it predicts observable fluxes of γ -rays and ultra-high energy cosmic rays from long strings stretching across the visible universe.

The OTW model requires the existence of a large-scale primordial magnetic field, with a strength $B_0 \approx 10^{-9}$ G at the present time, $t_0 \approx 10^{10}$ yr. For $t_{eq} < t < t_0$, where t_{eq} is the time of equal matter and radiation densities, the magnetic field is given by

$$B(t) \approx (t_0/t)^{4/3} B_0 \quad (1)$$

Superconducting strings move at relativistic velocities and develop electric currents as they cross the magnetic field lines. Long strings of length much greater than ct have the shape of random walks with a step $\sim ct$. In different horizon-size regions, the string crosses magnetic field lines in different directions, and the resulting current consists of harmonics of typical wavelength $\lambda \approx ct$. These harmonics oscillate with a typical timescale $\sim \lambda/c$, and are superimposed on higher frequency current waves generated at earlier times. The amplitude of the harmonic of wavelength λ is

$$i \approx e^2 \lambda B(\lambda) \approx 3 \times 10^{27} (t_0/\lambda)^{1/3} B_0 \text{ e.s.u. s}^{-1} \quad (2)$$

where e is the electron charge, $B_{-9} = B_0/10^{-9}$ G and I now use units in which $\hbar = c = 1$. For $\lambda < t_{eq}$, the amplitude of the current is independent of λ , $i \approx 3 \times 10^{29} B_{-9}$ e.s.u. s⁻¹. If strings are stretched by the expansion of the universe, then the wavelength λ grows proportionally to the string length L and the current amplitude decreases like L^{-1} . I shall disregard this effect here, since computer simulations indicate no substantial stretching of the strings (D. Bennett, personal communication).

An inhomogeneous current in a string results in charge accumulation, and portions of the string develop a charge per unit length comparable to the current, $\rho \approx i$. Electric and magnetic fields in the vicinity of the string are given by $E = 2\rho/r$ and $B = 2i/r$, where ρ and i are the local values of the linear charge density and current. In regions where $|\rho| > |i|$, the strong electric field near the string can produce particle-antiparticle pairs. Production of e^+e^- pairs begins when ρ becomes greater than $\sim m_e/e$, where m_e is the electron mass. If the string is positively charged, then electrons accumulate in bound states near the string and screen the electric field, so that $E < E_c \approx m_e^2/e$, for $r \gg m_e^{-1}$, while positrons are repelled from the string. Electrons cannot screen the field at distances $r \ll m_e^{-1}$, and pairs of heavier particles are produced when the charge density reaches the value $\rho \approx m/e$, where m is the particle mass. As a result the string develops layers of bound states; inner layers contain heavier particles. Heavy particles are stabilized by the binding energy, even though they may be unstable in vacuum. The masses of particles that can be produced by current harmonics of wavelength λ are limited by

$$m \leq ep \approx 10^{11} (t_0/\lambda)^{1/3} B_{-9} \text{ GeV} \quad (3)$$

for $\lambda > t_{eq}$ and $m \leq 10^{13} B_{-9}$ GeV for $\lambda < t_{eq}$.

The number of pairs produced per unit time per unit length

of string can be estimated as $dN/dt d\ell \approx \rho/\lambda$, and the energy output of the string is given by

$$(d\epsilon/dt d\ell) \approx \rho m_{\max}/\lambda \quad (4)$$

Here, $m_{\max}$ is the mass of the heaviest charged particle with $m \leq ep$.

The dissipation rate (4) should now be compared with the electromagnetic energy per unit length, $\rho^2 \ln(\lambda/\delta) \approx 100\rho^2$, where δ is the string thickness. (The energy of charge carriers, $\sim \rho^2/e^2$, is of the same order of magnitude¹.) The dissipation time for harmonics of wavelength λ is $\tau \approx 100\lambda\rho/m_{\max}$, and harmonics surviving at present have wavelengths $\lambda \geq 10^{-2} t_0 m_{\max}/\rho \approx 30(m_{\max}/\rho)$ Mpc. The energy output in pairs per unit length of string is mainly due to harmonics with the smallest λ , and using equations (2), (4) we find

$$\frac{d\epsilon}{dt d\ell} \approx 10^{44} B_{-9}^2 \kappa \text{ erg s}^{-1} \text{ Mpc}^{-1} \quad (5)$$

Here $\kappa = \min(B_{-9} m_{12}^{-1}, 3 \times 10^2)$ and $m_{12} = m_{\max}/10^{12}$ GeV.

The decay products of heavy particles produced by the strings reach the Earth in the form of γ -rays and cosmic rays. High-energy photons interact with the microwave background photons and lose their energy by e^+e^- pair production, and the pairs lose most of their energy by inverse Compton scattering. The resulting γ -ray spectrum peaks at $m_e^2/T \approx 10^5$ GeV, where T is the microwave background temperature⁸. (Here I assume that $m_{\max} > 10^5$ GeV. In the simplest grand unified models the 'desert' in the particle mass spectrum extends all the way from $\sim 10^2$ GeV to $\sim 10^{14}$ GeV. Then one expects to observe γ -rays with energies ~ 100 GeV and flux $F \approx 10^{-7} B_{-9}^2 \text{ cm}^{-2} \text{ s}^{-1}$.) For a string at a cosmological distance $\sim t_0$, the flux of 10^5 GeV photons on Earth is (per Mpc of string) $F \approx 10^{-15} \eta_\gamma \kappa B_{-9}^2 \text{ s}^{-1} \text{ cm}^{-2} \text{ Mpc}^{-1}$, where $\eta_\gamma \approx 1$ is the fraction of the energy of pairs converted into γ -rays. A detector with angular resolution $\sim 10^\circ$ pointed towards the string would detect a flux $F \approx 10^{-12} \eta_\gamma \kappa B_{-9}^2 \text{ cm}^{-2} \text{ s}^{-1}$. For $\eta_\gamma \kappa B_{-9}^2 \geq 10^{-2}$ this flux can be observed with existing detectors⁹, and the strings should actually be visible in the intensity distribution of 10^5 GeV γ -rays on the sky. A broad northern hemisphere survey has been performed¹⁰ using the Fly's Eye detector, but unfortunately it included only γ -rays with energies $> 10^6$ GeV. A similar survey with an energy cutoff below 10^5 GeV would give stringent constraints on the OTW model.

Equations (4) and (5) for the rate of particle production apply only to portions of string which do not move ultrarelativistically. However, oscillating loops of string tend to develop cusps where the string velocity momentarily reaches the speed of light. For a loop of length L , the ultrarelativistic motion with a Lorentz factor γ occurs during a time interval $t_\gamma \approx L/\gamma$ and is limited to the length of string $l_\gamma \approx L/\gamma^2$ (see ref. 11). The charge density in this segment of string is $\rho_\gamma \approx \gamma\rho$, the number of pairs produced to screen the electric field of the segment is $N_\gamma \approx \rho_\gamma l_\gamma/e$, and the total energy of pairs is $\Delta\epsilon \approx N_\gamma m_\gamma \gamma \approx \rho L m_\gamma/e$. Here, ρ is the typical charge density of the loop (in regions where $\gamma \approx 1$), m_γ is the mass of the heaviest charged particle with $m < ep_\gamma$, and I have assumed that $|\rho| > |i|$ near the cusp. The increase of the charge density towards the cusp is cut off at a critical value¹ $\rho_c \leq e\eta$, where η is the symmetry breaking scale of the string (the value of ρ_c is model-dependent). Hence we can write $m_\gamma \leq ep_c$ and $\Delta\epsilon \leq \rho\rho_c L$. For bosonic superconducting strings $\Delta\epsilon$ can be close to its maximum value, $\sim \rho\rho_c L$, which is comparable with the energy loss from the cusp in the form of low-frequency electromagnetic waves¹². Babul, Paczynski and Spergel have suggested⁶ that bursts of γ -rays from oscillating loops of string at high redshifts can be identified with the observed γ -ray bursters in the MeV range. Particle production by superconducting strings can provide a specific mechanism for their model.

In addition to γ -rays, the decay products of heavy particles produced by long strings can reach the Earth as high-energy protons and neutrinos with energies up to $\sim m_{\max}$ and flux

$F \approx 10^{-19} B_{-9}^2 \kappa m_{12}^{-1} \text{ cm}^{-2} \text{ s}^{-1}$ (assuming an angular resolution of $\sim 10^\circ$). Energies of the particles emitted from near-cusp regions can be as high as $\sim e\rho_c^2/\rho$, and can even exceed the string symmetry breaking scale, η . Predictions for the energy distribution and the flux of cosmic rays produced by the strings are more sensitive to the details of particle physics models than those for γ -rays, and we will not discuss them further in this paper.

I thank Drs J. W. Elbert, G. B. Field and J. E. Grindlay for helpful discussions. This work was supported in part by the National Science Foundation and by contributions from Denison Manufacturing Co. and General Electric Co.

Received 10 December 1987; accepted 22 January 1988.

1. Witten, E. *Nucl. Phys.* **B249**, 557–592 (1985).
2. Chudnovsky, E. M., Field, G. B., Spergel, D. N. & Vilenkin, A. *Phys. Rev. D* **34**, 944–950 (1986).
3. Ostriker, J. P., Thompson, C. & Witten, E. *Phys. Lett.* **B180**, 231–239 (1986).
4. Vilenkin, A. & Field, G. B. *Nature* **326**, 772–773 (1987).
5. Hill, C. T., Schramm, D. N. & Walker, T. P. *Phys. Rev. D* (in the press).
6. Babul, A., Paczynski, B. & Spergel, D. N. *Phys. Rev. D* (in the press).
7. Aryal, M., Vachaspati, T. & Vilenkin, A. *Phys. Lett.* **B194**, 25–29 (1987).
8. Hogan, C. J. & Rees, M. J. *Nature* **311**, 109–114 (1984).
9. Elbert, J. W. in *13th Texas Symposium on Relativistic Astrophysics* (ed. Ulmer, M. P.) 567–572 (World Scientific, Singapore, 1987).
10. Baltusaitis, R. M. *et al. Astrophys. J.* **297**, 145–150 (1985).
11. Vilenkin, A. in *300 Years of Gravitation* (eds Hawking, S. & Israel, W.) 499–523 (Cambridge University Press, 1987).
12. Vilenkin, A. & Vachaspati, T., *Phys. Rev. Lett.* **58**, 1041–1044 (1987).
13. Spergel, D. N., Piran, T. & Goodman, J. *Nucl. Phys. B* (in the press).

The nature and origin of interstellar diamond

David F. Blake*, Friedemann Freund*,
Kannan F. M. Krishnan†, Charles J. Echer†,
Ruth Shipp*, Theodore E. Bunch*,
Alexander G. Tielens*, Robert J. Lipari‡,
Crispin J. D. Hetherington† & Sherwood Chang*

* NASA Ames Research Center, Moffett Field, California 94035, USA

† National Center for Electron Microscopy,
Lawrence Berkeley Laboratory, Berkeley, California 94720, USA

‡ Surface Science Laboratories, Mountain View, California 94043, USA

Microscopic diamond was recently discovered¹ in oxidized acid residues from several carbonaceous chondrite meteorites (for example, the C δ component² of the Allende meteorite). Some of the reported properties of C δ seem in conflict with those expected of diamond^{1,3}. Here we present high spatial resolution analytical data which may help to explain such results. The C δ diamond is an extremely fine-grained (0.5–10 nm) single-phase material, but surface and interfacial carbon atoms, which may comprise as much as 25% of the total, impart an ‘amorphous’ character to some spectral data. These data support the proposed high-pressure conversion of amorphous carbon and graphite into diamonds due to grain–grain collisions in the interstellar medium⁴ although a low-pressure mechanism of formation cannot be ruled out.

The presence of krypton and nitrogen as well as isotopically anomalous xenon (‘CCFXe’, or ‘Xe-HL’) in the C δ residue^{5,6} has been interpreted to mean that the diamond is interstellar in origin^{1,2}. Several low-pressure^{1,7,8} as well as high-pressure⁴ mechanisms of formation for interstellar diamond have been discussed.

Fragments of the Allende CV3 meteorite were treated with acid and oxidant using a method similar to that in Lewis *et al.*¹. Preliminary observations revealed the presence of an apparent amorphous phase containing silicon and oxygen in addition to the diamond. The Si–O material was removed by further acid treatment. Diamond was identified by selected area electron diffraction. Diamond masses, 10–500 nm size, were observed.

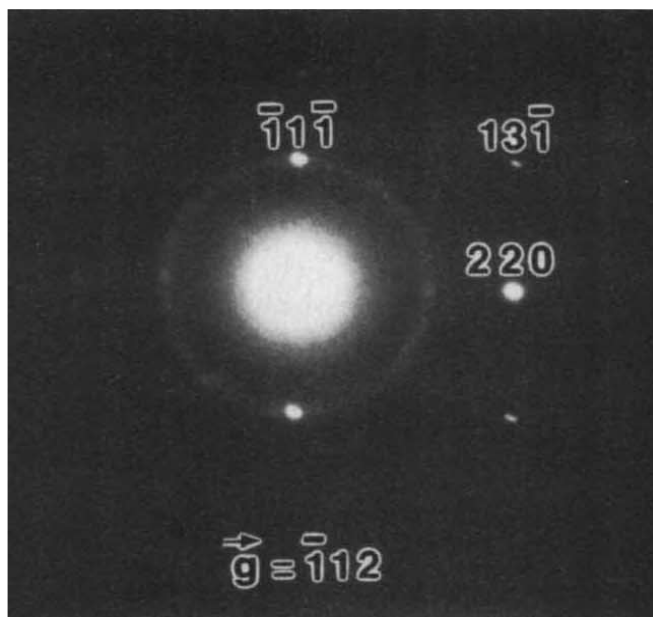


Fig. 1 Microdiffraction pattern from Allende C δ diamond. (001), (011), and (112) orientations were obtained; only the (112) direction is shown.

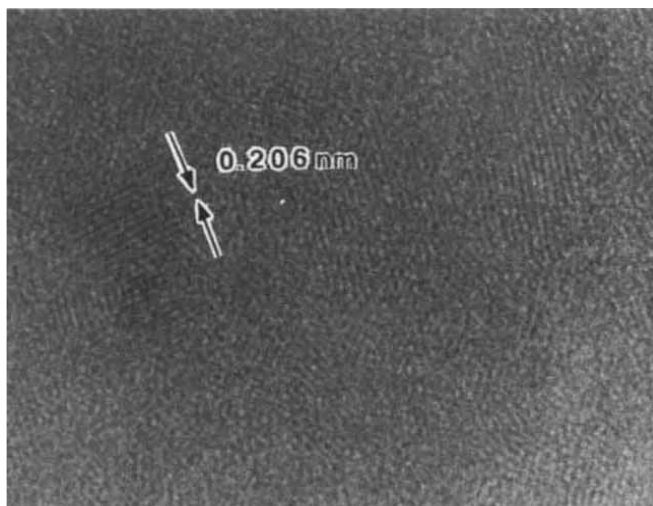


Fig. 2 High-resolution lattice image of (111) planes in a polycrystalline mass of C δ diamond. The microcrystallites are surrounded by apparently structureless regions and diamond crystallites in non-diffracting orientations.

Scanning electron microscopy of these masses revealed tightly packed microcrystallites and an apparent lack of porosity. Bright and dark-field transmission electron microscopy (TEM) of these masses revealed 0.5–10.0 nm crystallites with irregular morphology. Microdiffraction using a 20-nm stationary probe (Fig. 1) yielded a Debye–Scherrer powder pattern of diamond with superimposed single crystal spots, confirming the exceptionally small grain size of most crystals and the presence of a few larger individuals. The a parameter (the length of the cube edge) is 0.365 nm, about 2% larger than the published value of 0.35667 nm. This increase in the lattice constant is consistent with very small diamond crystallites, as Vanderbilt and Louie⁹ predict an increase in the C–C bond length by as much as 8.1% within the first two atomic layers of a free diamond surface. High-resolution TEM images of {111} planes of the C δ diamond (Fig. 2) do show 2–4 nm tightly joined crystallites separated

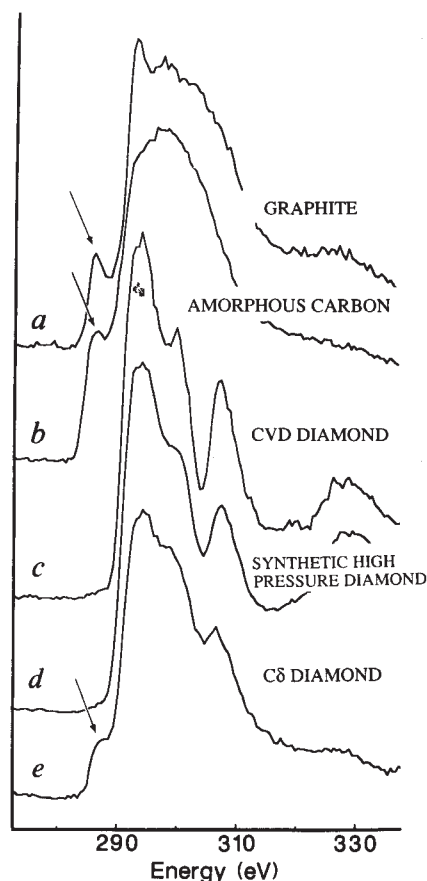


Fig. 3 EEL spectra for a number of carbon phases, showing the region in the vicinity of the carbon K edge. The region between 285 and 315 eV contains interpretable fine structure. Above 315 eV, multiple scattering (that is, thickness effects) dominates. *a*, Highly ordered pyrolytic graphite. *b*, Amorphous arc-sputtered carbon. *c*, Synthetic low-pressure CVD diamond (sample kindly provided by Dr R. Messier, Pennsylvania State University). *d*, Synthetic high pressure diamond. *e*, C δ diamond. Some chemical shifts in the carbon K edge are apparent between spectra. A significant pre-ionization edge structure is present in the graphite, amorphous carbon, and C δ diamond spectra (arrowed), representing carbon $1s \Rightarrow \pi^*$ transitions which are not allowed in the true diamond structure.

from one another by apparently structureless regions which may represent either crystallites in non-diffracting orientations, disordered interfacial regions, or both.

Analytical electron microscope microanalyses, using an energy-dispersive X-ray detector sensitive to all elements above boron in atomic number, showed that O and possibly N are the only impurities consistently present above 0.1–0.2%. The presence of substitutional N at 0.25% or less is supported by electron spectroscopy for chemical analysis (ESCA) and optical fluorescence studies on bulk C δ material. According to Carey *et al.*¹⁰ the H concentration measured by ion-probe analysis may exceed that of O by an order of magnitude.

Figure 3 shows electron-energy-loss spectra (EELS) from different forms of carbon. A comparison of the C K edge shape and fine structure showed C δ diamond masses to be largely diamond, but with a significant pre-edge absorption feature indicative of transitions of C $1s$ electrons into π^* orbitals. Such π^* states are absent in ideal sp^3 -bonded diamond but present in amorphous carbon, in graphite and at diamond surfaces as a result of restructuring^{9,11}.

Figure 4 shows that the ESCA spectrum of C δ exhibits a $\pi \Rightarrow \pi^*$ shake-up satellite arising from C $1s$ photoelectrons

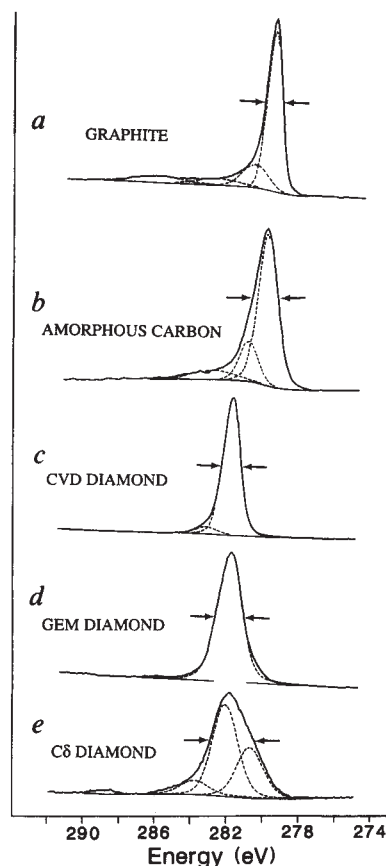


Fig. 4 ESCA spectra of the carbon $1s$ peak region of graphite, amorphous carbon and diamond materials (spectra aligned with reference to published carbon $1s$ -binding energies for graphite and diamond). Dotted lines indicate computer-generated gaussian peaks fitted to observed data. *a*, Highly ordered pyrolytic graphite. Note diffuse nature of $\pi \Rightarrow \pi^*$ shake-up satellites and asymmetry of measured peak. Peak width approaches instrumental resolution of 0.7 eV. *b*, Amorphous arc-sputtered carbon. Note marked asymmetry on high-energy side of peak. *c*, Synthetic low-pressure CVD diamond film. *d*, Gem-quality natural diamond. Note absence of $\pi \Rightarrow \pi^*$ shake-up satellite. Width slightly increased due to charging. *e*, C δ , showing a narrow shake-up satellite at 289 eV. The carbon $1s$ peak can be separated into two main components, probably due to carbon atoms in the bulk and in the interfacial regions.

which suffer an energy loss by exciting a π electron into a π^* orbital. Since purely sp^3 -bonded diamond lacks π states, the presence of such a shake-up peak may be an indication of restructured π -type C–C bonds at grain-grain interfaces (such as at the surface of diamond crystallites^{9,11}), or of sp^2 bonds in graphitic or 'amorphous' regions. The increased width of the C $1s$ peak is due to C atoms in slightly different environments. If one assumes a crystallite size of 2 nm, and if the restructuring affects three atom-layers, as many as 25% of all C atoms in C δ diamond would have electron states different from C atoms in the bulk. We therefore suggest on the basis of high resolution TEM, EELS and ESCA data, that the C δ diamond is single phase, but with a significant contribution from C atoms at and near tightly interlocked grain-grain interfaces which lend an apparent amorphous character to EELS and ESCA spectra.

Lewis *et al.*¹ suggest that the C δ diamond was formed at low pressure, by processes similar to those used in recent low-pressure chemical vapour deposition (CVD) laboratory syntheses¹². The CVD diamond we have characterized by electron microscopic techniques is comprised of large ($>1 \mu\text{m}$) euhedral crystals which are quite unlike C δ diamond. However, low-pressure diamond formation can occur under a wide variety of

conditions resulting in a range of microstructural features. We therefore cannot exclude a low pressure mechanism of formation. Rather, we cite a second possible mechanism of diamond formation in interstellar space. High velocity ($>10 \text{ km s}^{-1}$) grain-grain collisions behind supernova shock waves⁴ provide the high pressures required to convert amorphous or graphitic carbon grains into diamonds¹³. According to the grain-grain collision model, this conversion takes place by melting of carbon in the diamond stability field and extremely rapid crystallization from many nucleation sites. During crystallization, impurities that may be present in the original grain, for example krypton and xenon, adsorbed to amorphous or graphitic carbon, are expected to be excluded and to become trapped in the interfacial regions between diamond crystallites, while nitrogen may enter substitutional sites in the diamond structure. The prediction of the model regarding the grain size of the resulting diamonds is consistent with the TEM data showing the C δ diamond to consist of 0.5–10 nm, tightly joined, well crystallized microcrystallites. The cryptocrystalline nature of the C δ diamond masses, plus the apparent lack of porosity, strongly argue against agglomeration of pre-formed diamond crystallites through sintering.

The bulk of the carbon in primitive meteorites occurs as acid-insoluble, fine-grained particulate matter. The mechanisms for production of this material under solar nebula conditions are poorly understood¹⁴. It has previously been suggested that pre-solar origins can account for the minor fraction of the carbonaceous components that contain unusual isotopic ratios for H, C or N¹⁵. An interstellar origin for the C δ diamond,

which contains isotopically 'ordinary' carbon¹⁰, however, emphasizes the possibility that most of the reduced carbonaceous particles in primitive meteorites had a similar origin.

It remains to find diamond *in situ* in carbonaceous chondrite meteorites so that detailed petrographic investigations of textural relationships between diamond and other phases can be carried out. In addition, a further characterization of low pressure diamond is necessary to determine the range of structures and chemistries resulting from these non-equilibrium processes.

This research was sponsored in part by the Exobiology and Planetary Materials Programs of NASA, the NRC Resident Research Associateship Program (to D.B.), and DOE.

Received November 4 1987; accepted 18 February 1988.

1. Lewis, R. S., Tang, M., Wacker, J. F., Anders, E. & Steel, E. *Nature* **326**, 160–162 (1987).
2. Swart, P. K., Grady, M. M., Pillinger, C. T., Lewis, R. S. & Anders, E. *Science* **220**, 406–410 (1983).
3. Ash, R. D. *Meteoritics* **22**, (in the press).
4. Tielens, A. G. G. M., Seab, C. G., Hollenbach, D. J. & McKee, C. *Astrophys. J. Lett.* **319**, L103 (1987).
5. Ott, U., Mack, R. & Chang, S. *Geochim. cosmochim. Acta* **45**, 1751–1788 (1981).
6. Lewis, R. S., Anders, E., Wright, I. P., Norris, S. J. & Pillinger, C. T. *Nature* **305**, 767–771 (1983).
7. Hecht, J. H. *Nature* **328**, 765 (1987).
8. Nuth, J. A. *Nature* **329**, 589 (1987).
9. Vanderbilt, D. & Louie, S. G. *Phys. Rev. B* **30**, 6118–6130 (1984).
10. Carey, W., Zinner, E. & Lewis, R. *Meteoritics* **22**, (in the press).
11. Morar, J. F. *et al. Phys. Rev. B* **33**, 1346–1350 (1986).
12. Roy, R. *Nature* **325**, 17–18 (1987).
13. DeCarli, P. S. & Jamieson, J. C. *Science* **133**, 1821 (1961).
14. Kerridge, J. F., Shipp, R. & Chang, S. *Geochim. cosmochim. Acta* **51**, 2527–2540 (1987).
15. Lewis, R. S. & Anders, E. *Scient. Am.* **249**, 66–77 (1983).

Residual stress measurement by means of the thermoelastic effect

A. K. Wong, S. A. Dunn & J. G. Sparrow

Defence Science and Technology Organisation,
Aircraft Structures Division, Aeronautical Research Laboratory,
PO Box 4331, Melbourne, Victoria 3001, Australia

The term thermoelastic effect refers to the coupling between mechanical deformation and the change in thermal energy of an elastic material. The first theoretical treatment of this phenomenon is attributed to Lord Kelvin¹, and the resulting law states that the rate of change in temperature of a dynamically loaded body is directly related to the rate of change of the principal stress sum under adiabatic conditions. Although Kelvin's law has been well known for over a century, it is only in the past ten years that the thermoelastic effect has been exploited as a means for dynamic stress analysis. A system known as SPATE (stress pattern analysis by measurement of thermal emission) has been developed which can detect changes in infrared emission due to minute changes in the temperature of a dynamically stressed material. Recently it was discovered that the SPATE response or, more generally, the thermal response of a cyclically loaded body is not only a function of the dynamic part of the stress, but also of the static component². This finding has led to the suggestion that residual stresses within a material might be detected using this phenomenon, and here we present the first demonstration of such a means of residual stress measurement.

The experimental findings of Machin *et al.*² are important in two respects. The first, as mentioned above, raises the possibility of using the thermoelastic effect to measure residual stresses. The second raises the more fundamental question on the completeness of Kelvin's law as it does not account for the dependence of the temperature response on any static stress. In an attempt to resolve the latter problem, Wong *et al.*³ repeated the derivation of the thermoelastic theory from first principles, and found that the inclusion of higher-order terms can account for

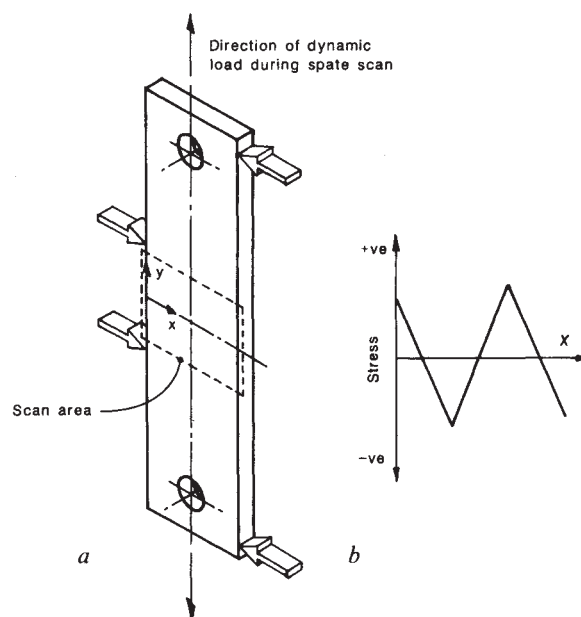


Fig. 1 Test specimen showing the four-point loading configuration (a) and a typical residual stress profile across the width of the specimen (b).

the stress-dependence of the thermoelastic effect. By treating the elastic moduli as general functions of temperature, it was shown that the mean stress-dependence of the thermoelastic 'constant' obtained in the experiments of Machin *et al.*² can be predicted with good accuracy. The revised theory further predicted that when a body is subjected to a single-frequency load, the temperature response should not only be at the loading frequency (as implied by Kelvin's law), but that there should also be a component at twice that frequency. In subsequent experiments, this predicted second harmonic response was indeed detected⁴, thereby strongly supporting the validity of the revised theory.

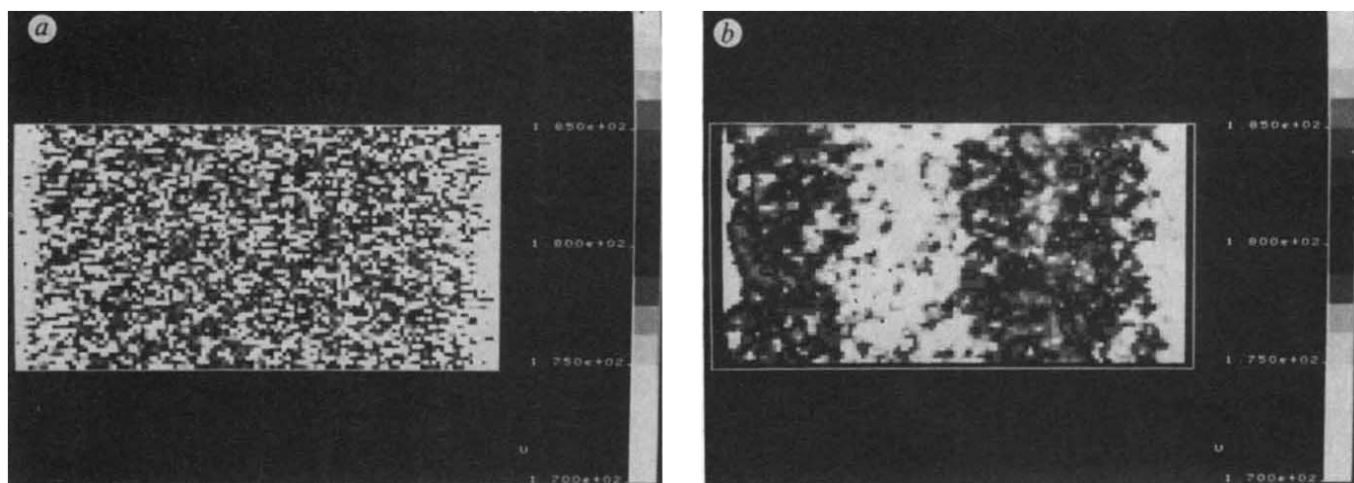


Fig. 2 Comparison of SPATE scans for the straight specimen (a), and the straightened specimen containing residual stresses (b).

The understanding of the reasons for the stress-dependence of the thermoelastic response now paves the way for exploiting this phenomenon for measuring residual stresses in a material. To demonstrate the feasibility of this technique, SPATE was used to scan several specimens containing a known distribution of residual stresses. Three specimens, of nominal dimensions 190 mm × 38 mm × 6.4 mm, were manufactured from a 2024 T351 aluminium alloy (yield stress of 325 MPa) which was assumed to be initially stress-free. The specimens were purposely machined curved, and residual stresses were then induced by straightening the specimens in a four-point loading rig (see Fig. 1a). For the specimens chosen, loads were selected to give maximum bending moments of ~1.5 kNm, which was sufficient to cause plastic deformation over part of the specimen's mid-section. Subsequent removal of the applied load results in a residual stress profile that is typically shaped as shown in Fig. 1b. A series of up to eleven strain gauges, equally spaced across the specimens, were used to record the strain profiles in the loaded and relaxed states. These strain data, together with the constitutive law in which any unloading process was assumed to be linearly elastic, enabled the stress profile to be determined. (Note that residual stresses may be determined from strain gauge data only if the strain gauges were bonded to the material in the stress-free state, and that the entire strain history is normally required.)

Each specimen was mounted in an axial loading rig and scanned while a 20 Hz cyclic load was applied in the longitudinal direction. The scans were performed on the sides opposite to where the strain gauges were mounted, and a matt-black paint was used to improve the infrared emissivity of these surfaces. A stress amplitude of 40 MPa (about a mean stress of 50 MPa) was selected as a good compromise between a dynamic stress level which was sufficiently high for obtaining a sizeable SPATE signal and low enough not to cause further plastic deformation. Scans of resolution 120 × 62 points were used to cover the full width of the specimens and over a length of ~20 mm of the central region where the residual stresses were expected to be essentially one-dimensional. An integration time of 0.8 s per scan-point was found to be sufficient, and the entire scan was completed in ~2 h. Due to alignment problems, some dynamic bending stresses were still present despite the specimens being essentially straight. But because such bending stresses varied linearly across the specimens and were easily detected (by means of a hand-held strain gauge), their effect on the scan data may be removed by the subtraction of this component.

For the purpose of comparison, a SPATE scan was obtained on a geometrically similar specimen which had not had any prior plastic deformation (and therefore contains no residual stresses). Figure 2 shows the comparison of the scans of the 'straight' and 'straightened' specimens. Although Kelvin's law would have predicted these two specimens to have the same thermal response, it is clear that the two SPATE scans are significantly different. Even though there were no stress concentration effects as the specimens were of uniform cross-section, the straightened specimens produced scans that showed four distinct bands. Because the presence of a compressive static stress results in a reduction in the thermal response whereas a tensile static stress has the opposite effect, these four bands reflect the two compressive and two tensile residual stress regions illustrated in Fig. 1b. As the scan data appeared to vary essentially in only one direction (that is, width-wise), 40 rows of the scan data were averaged to form a single line scan. Given that

$$\frac{1}{K_0} \frac{\partial K}{\partial s_m} = \lambda$$

where s_m is the total mean stress, K is the thermoelastic parameter and K_0 is the conventional thermoelastic constant, it may

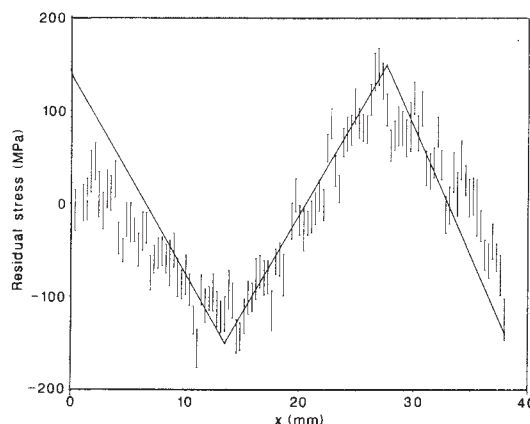


Fig. 3 Residual stress profile for a four-point straightened specimen. The vertical bars indicate the standard error for the average SPATE data, and the continuous line is the prediction based on the strain-gauge results.

be shown that for small changes in K

$$\frac{K - \bar{K}}{\bar{K}} = \frac{\lambda s_r}{(1 + \lambda s_a)}$$

in which s_r is the residual stress; s_a is the applied mean stress; and $\bar{K}$ is the mean value of K across the width of the specimen. Because

$$\frac{S - \bar{S}}{\bar{S}} = \frac{K - \bar{K}}{\bar{K}}$$

where S is the SPATE output signal and $\bar{S}$ is its mean value across the width of the specimen, the relationship between the residual stress and the SPATE output is given by

$$s_r = \frac{(1 + \lambda s_a)}{\lambda} \frac{S - \bar{S}}{\bar{S}}$$

Note that this procedure makes the evaluation of the residual stresses independent of the applied stress amplitude, as well as avoiding the necessity for determining K_0 .

For the material considered, a value of $\lambda = 3.02 \times 10^{-4} \text{ MPa}^{-1}$ (see ref. 3) was used; the results for one of the specimens are shown in Fig. 3. As can be seen, the residual stress profile derived from the SPATE data compares well with that obtained from the strain gauge results over most of the width of the specimen. The greatest deviation occurred on the left hand side of the figure, where the material had undergone compressive plastic deformation. For the three specimens considered, it was found that the residual stresses predicted by SPATE in this region were consistently below that of the strain gauge results. Similar differences have been reported in recent publications on acoustoelastic residual stress measurements^{5,6}, and the reason for such discrepancies could be attributed to the influence of compressive yielding on the physical properties of the material. On the other hand, the use of the simple constitutive law for converting the strain gauge data to stresses could well be inadequate for compressive plastic deformation, thus leading to errors in the residual stresses predicted by the strain gauges.

Our preliminary experiments furnish the first demonstration that residual stresses can be detected and measured by means of the thermoelastic effect. Much work is needed, however, before a practical instrument can be developed. In practice, the component under consideration would generally be more complex in shape so that the thermal pattern would have to be attributed to the distribution of both the dynamic stress and internal residual stress. To separate these two components, accurate measurements of the fundamental as well as of the second harmonic thermal response may be necessary⁴. This procedure could require an instrument much more sensitive than SPATE. Another area needing rigorous attention involves the effects of material texture. Effects such as those due to grain boundaries and crystallographic orientations are currently the main stumbling block for existing non-destructive residual stress measurement methods, such as X-ray diffraction and ultra-sonic techniques. Other techniques such as the Barkhausen noise analysis are unable to separate the effects of prior plastic deformation and residual stresses. Our experimental results have already raised some questions concerning the compressively yielded region, and a detailed experimental programme is currently under way to quantify this and other issues.

The authors thank Brian Quinn, Carlos Rey, Roy Bailey, Barry Ashcroft, Rodney Bateman, Ian Grundy and Silvia Gonzalez for their assistance.

Received 21 December 1987; accepted 8 February 1988.

1. Thompson, W. (Lord Kelvin) *Trans. R. Soc. Edinb.* **20**, 261-288 (1853).
2. Machin, A. S., Sparrow, J. G. & Stimson, M.-G. *Strain* **23**, 27-30 (1987).
3. Wong, A. K., Jones, R. & Sparrow, J. G. *J. Phys. Chem. Solids* **48**, 749-753 (1987).
4. Wong, A. K., Sparrow, J. G. & Dunn, S. A. *J. Phys. Chem. Solids* (in the press).
5. Fisher, M. J. & Herrmann, G. *Rev. Progr. Quant. NDE* **3B** 1283-1291 (1984).
6. Hirao, M. & Pao, Y.-H. *J. Acoust. Soc. Am.* **77**, 1659-1664 (1985).

Surface-plasmon microscopy

Benno Rothenhäusler* & Wolfgang Knoll†

* Physik-Department E22, Technische Universität München, James-Frank-Strasse, D-8046 Garching, FRG

† Max-Planck-Institut für Polymerforschung, Jakob-Welder-Weg 11, D-6500 Mainz, FRG

The imaging of low-contrast samples is a challenging task for optical measuring techniques, especially if high lateral resolution is also required. For example, a heterogeneously organized lipid monolayer transferred from the water surface to a solid substrate¹ still needs an additional contrast enhancement mechanism (the solubility difference for a fluorescing chromophore incorporated between the fluid and the crystalline domains of the monolayer) to be visualized by fluorescence microscopy. The mere thickness or index contrast between the different regions is not sufficient to use either phase contrast or Nomarsky microscopy² or the more recently developed Isoscope ellipsometer³. Here we describe a new microscope technique—surface plasmon microscopy (SPM)—which offers superior contrast without loss of spatial resolution by using plasmon surface polariton (PSP) fields instead of normal light as the illumination source. Such electromagnetic modes travel along a metal-dielectric interface as a bound, non-radiative surface wave, with its field amplitudes decaying exponentially perpendicular to the interface. Although photons can be converted into PSPs by means of a plasmon coupler (a grating or a prism in many cases) this 'light' differs considerably from plane electromagnetic waves⁴. PSPs are characterized by first, a pronounced dispersion (energy and momentum are not linearly related by the speed of light); and second, a field intensity that is concentrated at the interface and strongly enhanced there. Some of these properties make these modes a sensitive measure of interfaces and ultrathin films. If plasmon surface polariton fields are used to illuminate interfacial structures in light microscopy, high contrast without loss of spatial resolution can be obtained owing to the high sensitivity of the plasmon resonance coupling to (for example) small optical thickness variations of thin dielectric coatings.

The test samples we used for demonstration purposes were thin dielectric coatings with well-defined lateral structures on top of the silver film that carries the PSP field. These were fabricated by first evaporating 50 nm Ag on to the base of a glass prism, which was then coated by a multilayer assembly of four cadmium arachidate layers ($(\text{CH}_3-(\text{CH}_2)_{18}\text{COO}^-)_2\text{Cd}^{2+}$, CdA) by the Langmuir-Blodgett (LB) dipping technique⁵. Next, part of the multilayer assemblies was removed by photo-(oxidative) desorption with ultraviolet light through the regularly spaced holes of an electron microscope (EM) grid used as a mask⁶.

The schematic set-up of the SPM is shown in Fig. 1. The laser

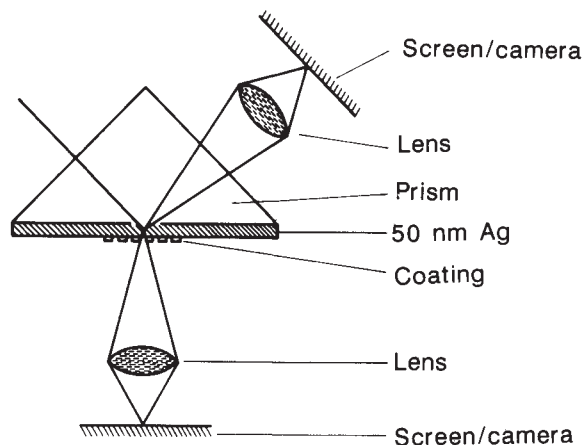


Fig. 1 Schematic of the optical set-up of SPM. Not to scale.

beam is coupled to the PSP modes at the silver-coating-air interface via a prism in the usual Kretschmann configuration⁷. Resonant excitation of PSP in the laterally heterogeneous interface occurs wherever the momentum matching condition $k_{\text{photon}} = k_{\text{sp}}$ between the parallel component of the photon wavevector, k_{photon} , and the PSP wavevector, k_{sp} , is fulfilled⁸. The latter depends on the optical thickness of the coating which results in a thickness-dependent resonance angle for PSP excitation⁹.

Those areas which carry a PSP field couple out some of the intensity to the air side, due to the intrinsic roughness of the interface. The scattered light can be viewed with an eyepiece of sufficient magnification. This mode of operation thus has some common features with dark-field illumination in ordinary microscopy. A photograph of such a partially radiating test structure is shown in Fig. 2. The EM grid used as a mask for the ultraviolet desorption process of the four deposited LB layers was a crossed grating. The grid structure (diameter 3 mm) was not illuminated completely and homogeneously because the laser beam was neither expanded nor spatially filtered. The resonance angle θ_1 was chosen to excite PSP in the LB film-coated areas (for example, the rim of the grid). Consequently only those regions couple out the PSP light, while the bare silver areas remain dark.

More effectively, and even in the absence of roughness, PSP couple out via the prism (see Fig. 1). Again this scattered light can be Fourier-backconverted into real space by means of a simple lens, to form an image of the interface. Figure 3 shows the result for two different angles of incidence of the exciting laser beam. For this example, the imaged structure was produced by ultraviolet desorption through a slotted grid. The intensity distribution presented in Fig. 3a is obtained when PSP are excited in the bare silver regions at $\theta_0 = 46.4^\circ$. These then appear dark because most of the electromagnetic field energy fed into the PSP is dissipated into heat and only a little light is reradiated into the prism. The LB film-coated areas, on the other hand, are not at resonance so all the excitation laser light is totally reflected.

The intensity distribution can be fully reversed by changing the angle of incidence to $\theta_1 = 53.6^\circ$. Now the coated areas are at resonance and emit little light, and vice versa (Fig. 3b).

This contrast mechanism is illustrated schematically in Fig. 4. Plane waves that impinge upon the interface at an angle θ_0 (Fig. 4a) excite PSP only in those areas that fulfil the resonance condition. Their corresponding reflectivity curve is given in Fig. 4b, full line. Areas with a different coating thickness (right-hand side of Fig. 4a) are characterized by a reflectivity curve which is slightly shifted in angular position (dashed line in Fig. 4b) with a shifted resonance angle, θ_1 . The reflectivity of these regions at θ_0 is nearly unity so that they appear bright in the reflected image.

The intensity contrast that can be achieved between areas of different coating thicknesses hence is closely related both to the width of the PSP resonance and to the relative shift of the resonance angle that any coating produces as a result of the modified dispersion curve for PSP.

It has been shown that a single lipid monolayer of thickness $d \approx 2.5$ nm produces, with laser light of wavelength $\lambda = 633$ nm a resonance shift of $\Delta\theta_s \approx 0.5^\circ$ (ref. 9). This must be compared with the width of the resonance which for silver is also about $\Delta\theta_w \approx 0.5^\circ$ (ref. 10). If we define the contrast between two areas as $K = (I_{\text{max}} - I_{\text{min}})/(I_{\text{max}} + I_{\text{min}})$ it follows that with $I_{\text{max}} \approx I_0$, the incident laser intensity, and $I_{\text{min}} \approx 0$ (see Fig. 4b) a bilayer only 5 nm thick yields maximum contrast $K = 1$ relative to bare silver areas. A further thickness increase does not have any additional effect, which means that SPM is particularly sensitive to small variations in thickness. Maximum contrast resolution, $\partial I/\partial d$, is obtained for the thinnest coatings and amounts to about $\partial I/\partial d = 0.4 I_0 \text{ nm}^{-1}$. This estimate is based on the evaluation of the maximum slope, $(dR/d\theta)_{\text{max}}$, of the angular reflectivity curve.

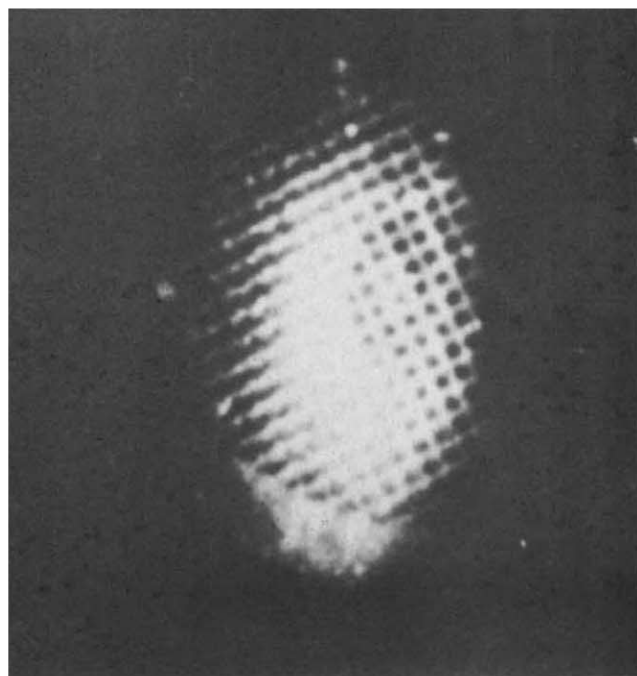


Fig. 2 Image of a test structure illuminated by PSP modes. The bright areas carry the plasmon field, which is partly coupled out to the air side through intrinsic roughness.

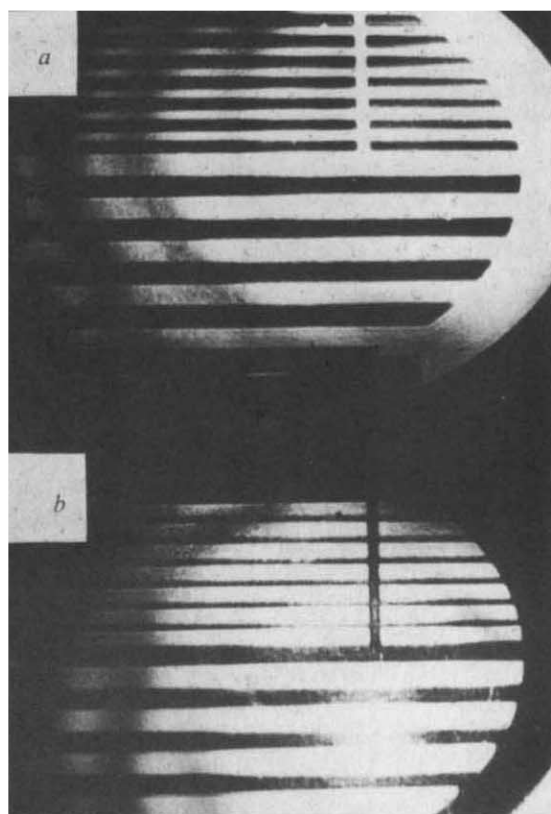


Fig. 3 Image of a test structure deposited onto the PSP carrying silver film, viewed from the prism side. Those areas which are at resonance for PSP generation (bare silver in a and the LB film-coated regions in b) appear dark because most of the plasmon field energy is dissipated into heat, while the other regions still fulfill the condition for total internal reflection and hence appear bright. By changing the angle of incidence, the intensity distribution can be reversed.

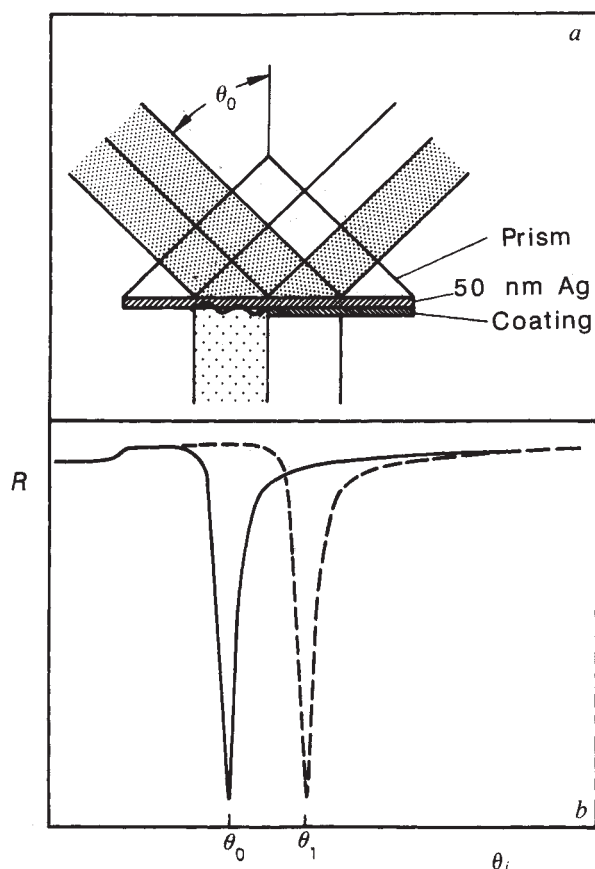


Fig. 4 *a*, Schematic illustration of the contrast mechanism in SPM. Half the silver film is covered by a thin coating which shifts the resonance condition for plasmon coupling. At θ_0 only in the bare silver areas PSP (wavy arrow) are excited. A small fraction of this light is coupled out to the air side, while almost nothing is reradiated to the prism side. The coated regions are off resonance so that all laser light is totally reflected. *b*, Reflectivity, R versus angle of incidence, θ_i , for bare silver (full line) and for coated regions (dashed line).

tivity curve and measured values for the resonance-shift by LB layer coatings.

We suggest that this high sensitivity is because we are dealing with a resonance contrast whereas all other microscope techniques capable of resolving thickness variations in the 10-nm range are, in one way or another, based on interference contrast. Interference curves are, however, much smoother functions than resonance curves.

As for any microscopy, contrast problems in SPM can not be discussed without addressing the question of lateral resolution. Because the detection scheme contains plane-wave optical components, SPM is also limited by diffraction—lateral resolution is in the 1- μm range. If PSP waves are used for the illumination of the object under investigation, additional effects that could reduce the lateral resolution must be considered.

PSPs are propagating modes^{8,11} which are strongly damped in their propagation direction due to intrinsic dissipation^{13,14} and radiative damping^{8,15}. A quantitative measure of how far a PSP mode travels is the propagation length L_x , which is defined by the $1/e$ attenuation of the PSP field in the propagation direction, and is directly related to the imaginary part of the complex PSP wave vector, k_{ix} , by $L_x = (2k_{ix})^{-1}$. If radiative damping can be neglected, L_x is predominantly determined by the complex dielectric function of the metal. As a consequence, L_x is sensitively wavelength-dependent: between $\lambda = 450$ and 1,150 nm, L_x varies for silver by a factor of 50, from 4.4 to 200 μm . L_x , however, determines to what extent PSP modes average over lateral heterogeneities¹⁶. This means that high

spatial resolution in SPM is achieved only for lossy PSP modes, for example with silver only in the ultraviolet spectral range. With gold, however, even for $\lambda \geq 520$ nm, where the onset of the interband transition increases the intrinsic damping considerably, $L_x \leq 1 \mu\text{m}$ can be achieved¹⁷.

The countercurrent effects of L_x on spatial resolution and contrast are quite obvious, because decreasing L_x means, from the above, an increase of k_{ix} , and thus an increase of the width of the PSP resonance. As demonstrated in Fig. 4, however, this reduces the contrast between different areas.

Finally, the examples presented above were all based on the contrast between silver areas with LB film coatings of different thicknesses. However, any difference in the optical properties of two materials that manifest themselves in different dispersion curves for PSP can be used to construct an image. In particular one might envisage (1) different metals that carry the PSP field; (2) different indices of refraction (both real and imaginary part) of equally thick coatings; or (3) roughness contrast.

We have restricted the examples to PSP excitation by a prism in the Kretschmann configuration although the Otto set-up, as well as a hybrid configuration (B.R. & W.K., in preparation) or even a grating coupling are feasible. Experiments along these lines are currently under way.

We acknowledge financial support by the Deutsche Forschungsgemeinschaft and by the Leonard-Lorenz-Stiftung, München.

Received 5 October 1987; accepted 5 January 1988.

1. Lösche, M. *et al. Thin Solid Films* **117**, 269–280 (1984).
2. Krug, W., Rienitz, J. & Schulz, G. *Beiträge zur Interferenzmikroskopie* (Akademie, Berlin, 1961).
3. Stenberg, M., Sandström, T. & Stibler, L. *Mat. Sci. Engng.* **42**, 65–69 (1980).
4. Burstein, E., Chen, W. P., Chen, Y. J. & Hartstein, A. *J. Vacuum Sci. Tech.* **11**, 1004–1024 (1974).
5. Kuhn, H., Möbius, D. & Bücher, H. in *Physical Methods of Chemistry* (eds Weissberger, A. & Rossiter, B. W.) pt IIB, Ch. VII (Wiley, New York, 1972).
6. Rothenhäusler, B. & Knoll, W. *Appl. Phys. Lett.* **51**, 783–785 (1987).
7. Kretschmann, E. *Opt. Commun.* **6**, 185–187 (1972).
8. Raether, H. in *Physics of Thin Films* (eds Hass, G., Francombe, M. H. & Hoffmann, R. W.) Vol. 9, 145–261 (Academic, New York, 1977).
9. Gordon, J. G., II & Swalen, J. D. *Opt. Commun.* **22**, 374–376 (1977).
10. Rothenhäusler, B. & Knoll, W. *Opt. Commun.* **63**, 301–304 (1987).
11. Otto, A. *Optik* **38**, 566–577 (1973).
12. Schoenwald, J., Burstein, E. & Elson, J. M. *Solid State Commun.* **12**, 185–189 (1973).
13. Inagaki, T., Kagami, K. & Arakawa, E. T. *Phys. Rev.* **B24**, 3644–3646 (1981).
14. Rothenhäusler, B., Rabe, J., Korpiun, P. & Knoll, W., *Surface Sci.* **137**, 373–383 (1984).
15. Hornauer, D.-L. *Opt. Commun.* **16**, 76–79 (1976).
16. Rothenhäusler, B. & Knoll, W., *Surface Sci.* **131**, 585–594 (1987).
17. Alexander, R. W., Jr *et al. J. chem. Phys.* **59**, 3492–3494 (1973).

Direct imaging of an adsorbed layer by high-resolution electron microscopy

D. A. Jefferson & P. J. F. Harris*

Department of Physical Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EP, UK

The catalytic activity of metal surfaces can be strongly modified, either detrimentally or beneficially, by the presence of pre-adsorbed material. In understanding these changes in activity it is important to elucidate the structure of the adsorbed layer and its relation to the metal surface. Traditionally this has been achieved by low-energy electron diffraction (LEED), but this requires large single-crystal specimens, which bear no resemblance to real catalysts, and the results are often difficult to interpret. We show here that high-resolution electron microscopy (HREM) can be used to obtain direct images of ordered layers of adsorbed sulphur on the surfaces of very small platinum particles in specimens that closely resemble commercial catalysts. Images of this

* Present address: T.I. Research, Hinxton, Saffron Walden, Essex CB10 1RH, UK

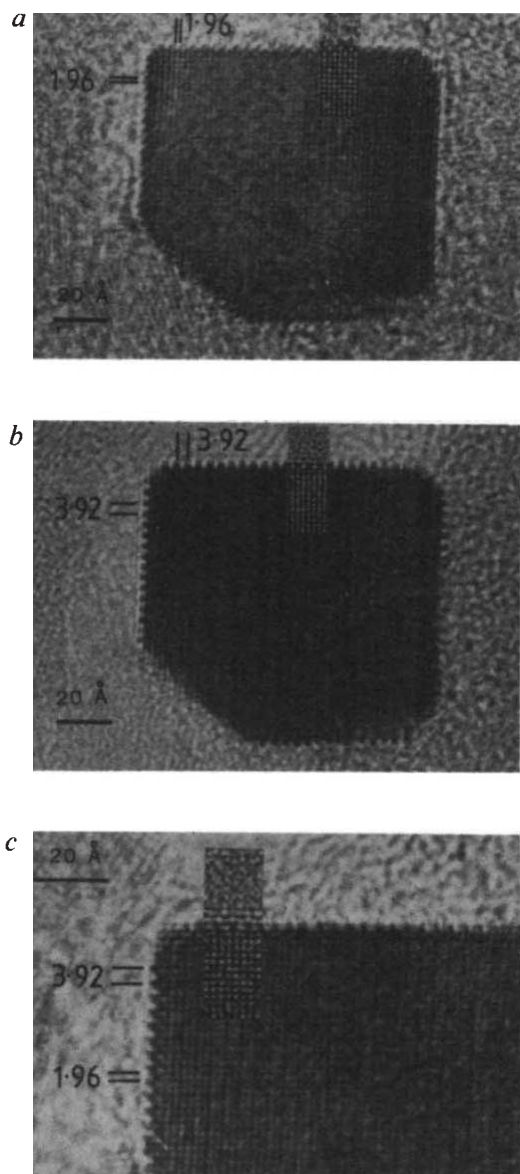


Fig. 1 HREM images of an approximately cubic particle of sulphided platinum, viewed down [100]. *a*, Objective lens defocus of approximately -550 Å; *b*, -250 Å; *c*, -150 Å. Defocus values were determined from the computer-simulated model images discussed in the text, which are shown inset for each. The surface periodicity in *b* and *c* (in ångströms) is indicated.

kind could lead to a more detailed understanding of the interaction of adsorbates with real metal catalysts than has hitherto been possible.

The work described here represents an extension of previous studies which have shown, by conventional bright-field transmission electron microscopy, that heating platinum/alumina specimens in sulphur-containing atmospheres induces (100) faceting of the platinum particles^{1,2}. Consequently, single-crystal particles in sulphur-treated specimens often exhibit approximately cubic shapes. This study involved imaging these faceted particles at much higher resolution with a modified JEM-200CX electron microscope developed for the study of ultrafine catalyst particles^{3,4}, operated at 200 kV. The specimens examined comprised metallic platinum particles of mean size ~ 113 Å dispersed on a self-supporting alumina film^{5,6}. Sulphur deposition was done by heating to 500 °C in a flowing mixture containing 100 p.p.m. of hydrogen sulphide in hydrogen^{1,2}.

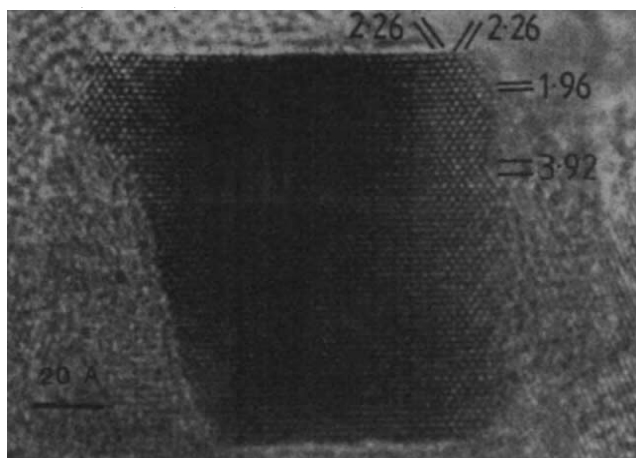


Fig. 2 HREM image of a similar particle to that in Fig. 1, but with one cube edge partly eroded, viewed down [110]. The relevant periodicities (Å) are indicated.

The objective lens characteristics of the microscope used (C_s , 0.52 mm; C_c , 1.05 mm) were such that at a focus slightly beyond the optimum or 'Scherzer' position⁷, the interpretable point resolution extended out to 1.95 Å with an absolute information limit of 1.7 Å. As in previous studies of surface restructuring of metals and oxides⁸⁻¹⁰, surface detail was observed in profile at the vertical edges of the crystallites, but in this case evidence of an apparent surface superlattice on the upper and lower faces was also noted. Micrographs were recorded at electron-optical magnifications of $\sim 475,000$ and 700,000, astigmatism being corrected by observing the granularity of amorphous regions of the specimen support, and particular care was taken to avoid inclination of the incident beam with its attendant difficulties in image interpretation¹¹. Observation of the images at greatly enhanced magnification by a television pick-up system was used to ensure that there were no changes in image detail arising from electron irradiation, and no observable carbon contamination was noted.

Typical micrographs of an approximately cubic crystallite of platinum (dimensions perpendicular to the incident beam of 104 Å $\times$ 92 Å) are shown in Fig. 1. Figure 1*a* shows an image of the crystal at approximately the optimum focus, with the bulk region showing only the crossed (200) fringes of 1.96 Å spacing, as expected for the [100] orientation of platinum metal, but the bright fringe at the edges of the crystal is clearly of complicated structure, with some indications of a doubling of the repeat distance in the direction parallel to the edge. At a position closer to the gaussian focus (Fig. 1*b*) this doubled repeat (3.92 Å) is clearly evident: furthermore, an apparent doubling of the lattice repeat within the crystal can now be observed. Both features are also evident in the enlarged image of Fig. 1*c*: in this case the focus position is slightly beyond the optimum one and the image contrast is reversed, with the platinum atoms appearing as white dots.

In addition to the [100] projection, images could also be recorded from particles in other directions, a [110] image being shown in Fig. 2. In this case the principal contrast within the crystal comes from two sets of 1.96 Å (200) fringes: the latter are parallel to the straight edges of the crystal. The crystal thickness clearly increases towards a line running down the centre: given the general near-cubic morphology it seems highly probable that Fig. 2 shows a micrograph of a roughly cubic crystal with the electron beam perpendicular to one of the cube edges. In the image of Fig. 2, the bright fringe on the upper and lower (100) faces appears to be perfectly normal, with no increased periodicity, but an apparent doubling of the (200)

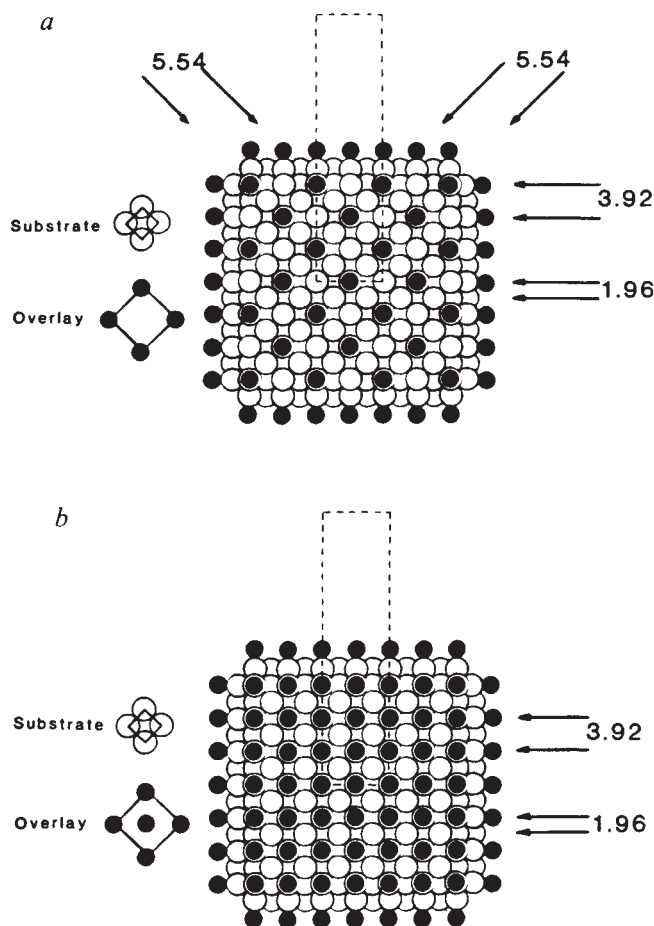


Fig. 3 Diagrammatic arrangements of the two possible sulphur overlays on a small cubic crystal of platinum, viewed down [100]. *a*, $p(2 \times 2)$ overlay; *b*, $c(2 \times 2)$. In both cases the unit cell used for computer image-simulation and the relevant periodicities are indicated.

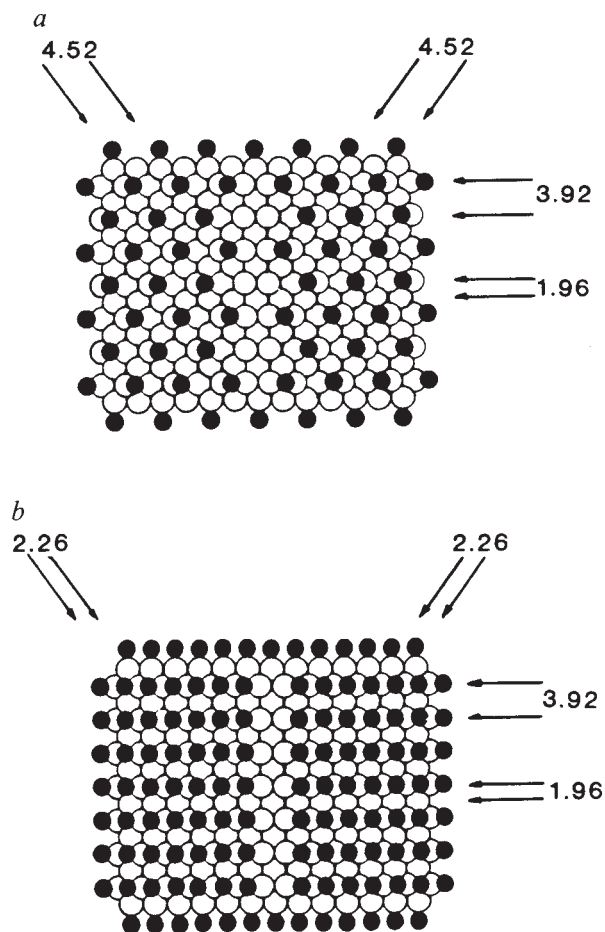


Fig. 4 As for Fig. 3, but with the crystals viewed down the [110] direction.

fringes is observed within the crystal bulk, particularly in the thicker regions.

The surface features observed in these images can arise in three ways; by formation of a surface sulphide, by restructuring of the platinum (100) surface or by the formation of an adsorbed layer of sulphur atoms. As regards the first, the conditions of preparation used were not favourable for the formation of bulk sulphide^{12,13}; furthermore, the changes in lattice periodicity expected for an epitaxial sulphide growth were not observed, even though recent work on the growth of Cu_2O on copper¹⁴ has indicated that even a single *complete* layer of oxide can be observed. For the second alternative, there is no LEED evidence to suggest that any restructuring of the platinum (100) face occurs on exposure to sulphur. In addition, restructuring of these faces to produce a doubling of the repeat length at the surface would imply a 'corrugated' (100) surface composed of tiny (110) facets, and although trial image simulations suggest that this could give contrast effects similar to those observed (indeed a surface arrangement of any atom could, at suitable thicknesses, give this contrast), it would seem very unlikely, as the proven effect of sulphur exposure is to eliminate all faces other than (100). Finally, energy-dispersive X-ray emission analysis upon groups of particles definitely indicates the presence of sulphur: consequently we must therefore consider the third alternative, namely that we are observing directly a layer of sulphur adatoms.

To construct a model of adsorbed sulphur which accounts for the observed image contrast, two main factors must be taken into account. Firstly, any arrangement must be compatible with existing LEED studies on macroscopic crystals^{15,16}, which indicate two possible types of overlays, either primitive or centred 2×2 arrays, corresponding to coverages of 25 and 50%, respectively. The arrangements of sulphur atoms are shown diagrammatically in Figs 3 and 4. When viewed in [100] projection (Fig. 3) it can be seen that both models will produce a doubling of the local repeat on the side faces of the cube, although the $p(2 \times 2)$ case might be expected to give a superlattice of 5.54 Å fringes at 45° to the cube edges on the upper and lower faces of the crystal: this was not observed in practice. In [110] orientation (Fig. 4) the $p(2 \times 2)$ model would give rise to a doubled repeat on the vertical faces, not seen in the real images, whereas the centred overlay would be quite compatible with the experimental observations. Argument on purely geometric grounds, therefore, would favour the $c(2 \times 2)$ adsorbate structure.

Although the geometrical considerations discussed above imply that the image contrast arises from the $c(2 \times 2)$ adsorbate layer, the question of whether or not adsorbed sulphurs can produce sufficient contrast to be observed, especially when viewed against a background of supporting Al_2O_3 , remains to be answered. To investigate this, a full set of multislice image simulations^{17,18} were done with the quasi-periodic cell indicated

in Fig. 3, with dimensions of $7.82 \text{ \AA} \times 31.32 \text{ \AA}$, and a sampling interval of $0.06 \text{ \AA}$, corresponding to an array size of 512×128 potential points. Owing to the extremely strong scattering of the platinum atoms, the slice thickness was $0.654 \text{ \AA}$ such that each platinum atom was essentially regarded as three scattering centres in a direction parallel to the incident electron beam, and there were three different types of slice, one of which contained only sulphur atoms.

The computations were performed to a crystal thickness of $100 \text{ \AA}$, against a $40 \text{ \AA}$ thick background of amorphous Al_2O_3 . Simulation of the contrast from the latter was accomplished by the method of Gai *et al.*¹⁹, taking random coordinates for aluminium and oxygen atoms within each slice computed. For this background, the slice thickness used was $2.5 \text{ \AA}$. At one particular defect of focus, all the relevant factors affecting image contrast, namely spherical aberration ($C_s = 0.52 \text{ mm}$) focus spread ($83 \text{ \AA}$) and beam divergence (10^{-3} mrad) were included. Owing to the excessive demands on computational resources required by the programs used²⁰ for this full calculation, however, subsequent simulations were performed without including the last two factors, where it was confirmed that their effect was merely to diminish the overall magnitude of the contrast observed. In such cases the resolution was artificially limited to $1.8 \text{ \AA}$.

The results of these simulations are shown as insets on Fig. 1. They confirm that an adsorbed sulphur overlay can indeed produce enough contrast to be visible in the experimental image, and tests with varying thicknesses of amorphous background suggest that this would be valid for the $c(2 \times 2)$ overlay until the background medium was at least $120 \text{ \AA}$ thick. For the $p(2 \times 2)$ overlay, where the density of sulphur atoms in projection was only 50% of that for $c(2 \times 2)$, the overlay was almost invisible when the background thickness reached $50 \text{ \AA}$: consequently we must conclude that the observed images correspond to the latter structure.

Although the results described here refer to an ordered overlay, the use of the HREM technique would allow true imaging of a non-periodic arrangement, if this were present. Exactly how much structural information is obtainable remains to be seen, but in principle, by using multislice procedures, it may be possible to determine the exact location of adsorbate atoms and their spatial relation to those of the support, as the computation involved, although considerable, is less than that required for complete elucidation of a LEED pattern. In such cases we would then have an exceptionally powerful technique for determining the structure of adsorbates on real catalysts.

We thank the SERC for providing electron-microscopic facilities, and the Harwell Laboratory for the provision of materials. D.A.J. is indebted to Dr J. C. Barry, Arizona State University for assistance in the multislice computations. Part of this work was done under a NSF program.

Received 7 December 1987; accepted 16 February 1988.

- Harris, P. J. F. *Nature* **323**, 792–794 (1986).
- Harris, P. J. F. *Surface Sci.* **185**, L459–L466 (1987).
- Jefferson, D. A. *et al.* *Nature* **323**, 428–431 (1986).
- Duff, D. G., Curtiss, A. C., Edwards, P. P., Jefferson, D. A. & Johnson, B. F. G. *Angew. Chem.* **26**, 676–678 (1988).
- Harris, P. J. F., Boyes, E. D. & Cairns, J. A. *J. Catal.* **82**, 127–146 (1983).
- Harris, P. J. F. *J. Catal.* **97**, 527–542 (1986).
- Erickson, H. P. & Klug, A. *Phil. Trans. R. Soc. B* **261**, 105–118 (1971).
- Marks, L. D. & Smith, D. J. *Nature* **303**, 316–318 (1983).
- Warble, C. E. *Ultramicroscopy* **15**, 301–309 (1984).
- Smith, D. J., Bursill, L. A. & Jefferson, D. A. *Surface Sci.* **175**, 673–683 (1986).
- Smith, D. J., Saxton, W. O., O'Keefe, M. A., Wood, G. J. & Stobbs, W. M. *Ultramicroscopy* **11**, 263–282 (1983).
- Oudar, J. *Catal. Rev. Sci. Eng.* **22**, 171–195 (1980).
- Bartholomew, C. H., Agrawal, P. K. & Katzer, J. R. *Adv. Catal.* **31**, 135–242 (1982).
- Curtiss, A. C. *et al.* *J. phys. Chem.* (in the press).
- Heegemann, W., Meister, K. H., Bechtold, E. & Hayek, K. *Surface Sci.* **49**, 161–180 (1975).
- Fischer, T. E. & Kelemen, S. R. *Surface Sci.* **69**, 1–22 (1977).
- Cowley, J. M. & Moodie, A. F. *Acta crystallogr.* **A10**, 609–619 (1957).
- Goodman, P. & Moodie, A. F. *Acta crystallogr.* **A30**, 280–290 (1974).
- Gai, P. L., Goringe, M. J. & Barry, J. C. *J. Microsc.* **142**, 9–24 (1986).
- Ishizuka, K. thesis, Arizona State Univ. (1986).

Electron microscopy on the $T_c = 110 \text{ K}$ (midpoint) phase in the system $\text{Bi}_2\text{O}_3\text{--SrO--CaO--CuO}$

H. W. Zandbergen*, Y. K. Huang†, M. J. V. Menken†, J. N. Li, K. Kadowaki†, A. A. Menovsky†, G. van Tendeloo‡ & S. Amelinckx‡

*Gorlaeus Laboratories, State University Leiden, PO Box 9502, 2300 RA Leiden, The Netherlands

†Natuurkundig Laboratorium der Universiteit van Amsterdam, Valckenierstraat 65, 1018 XB Amsterdam, The Netherlands

‡Center for Electron Microscopy, University of Antwerp, RUCA, Groenenborgerlaan 171, Antwerp, Belgium

The system is known to have at least two phases exhibiting high superconducting transition temperatures (T_c) of 85 and 110 K. It is well established that $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$ has a T_c of $\sim 85 \text{ K}$. To date it is not known which phase in the system exhibits $T_c = 110 \text{ K}$. We performed high-resolution electron microscopy and analytical electron microscopy on material having a nominal composition $\text{BiSrCaCu}_2\text{O}_x$, and containing a relatively high fraction of the phase with $T_c = 110 \text{ K}$. Two structures related to the structure of $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$ are found. These structures and their possible compositions are discussed.

Since the discovery of ceramic materials exhibiting high superconducting transition temperatures, much research on these materials has been carried out. One of the goals of the research was to discover materials with a T_c higher than that of the widely studied compound $\text{YBa}_2\text{Cu}_3\text{O}_7$. Because the high T_c is believed to be associated with the two-dimensional nature of $\text{YBa}_2\text{Cu}_3\text{O}_7$, one of the aims of the research was to try to substitute copper in known layered compounds.

One of the layered compounds resembling the perovskite-like structure of $\text{YBa}_2\text{Cu}_3\text{O}_7$ is $\text{Bi}_4\text{Ti}_3\text{O}_{12}$ ¹. The structure of this compound is composed of a perovskite block of three units stacked with a Bi_2O_2 layer. A large variety of phases exists, all having that Bi_2O_2 layer, but with a different number of units in the perovskite block. Michel *et al.*² studied the superconducting behaviour of a compound, which they reported to be $\text{Bi}_2\text{Sr}_2\text{Cu}_2\text{O}_{7+\delta}$ with a T_c of 20 K.

Maeda *et al.*³ investigated the system $\text{CaO--SrO--Bi}_2\text{O}_3\text{--CuO}$ and found a new superconductor with $T_c = 110 \text{ K}$, and claimed the composition to be $\text{BiSrCaCu}_2\text{O}_x$.

Research at many laboratories was immediately focused on this new phase. Many researchers were readily able to synthesize a compound with $T_c \sim 85 \text{ K}$, but were unable to reproduce the result of Maeda *et al.* The elucidation of the composition and the structure of the 85-K phase took several weeks, suggesting that this material is much more complex than $\text{YBa}_2\text{Cu}_3\text{O}_7$. The structure is found to be very complicated with a long period modulation. The basic structure, however, has been determined by several groups⁴, with an approximate composition of $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$.

One of the problems in determining the composition of the compound with T_c of 85 K is the composition range of the material. This is due partly to the existence of a region of solid solution (H. W. Z. *et al.*, manuscript in preparation) with a varying Ca:Sr ratio, but also because defects are introduced⁵ in the structure, changing the composition.

Apart from the modulation, a consensus has been obtained on the substructure and its superconducting behaviour. The 110-K material is still unresolved. Here we report, this phase was only a minority phase, making it particularly difficult to elucidate its structure and composition.

Here we report an electron microscopy study on a nominal mixture of $\text{BiSrCaCu}_2\text{O}_x$ exhibiting almost zero resistance and a considerable drop in the magnetization at 107 K. We show

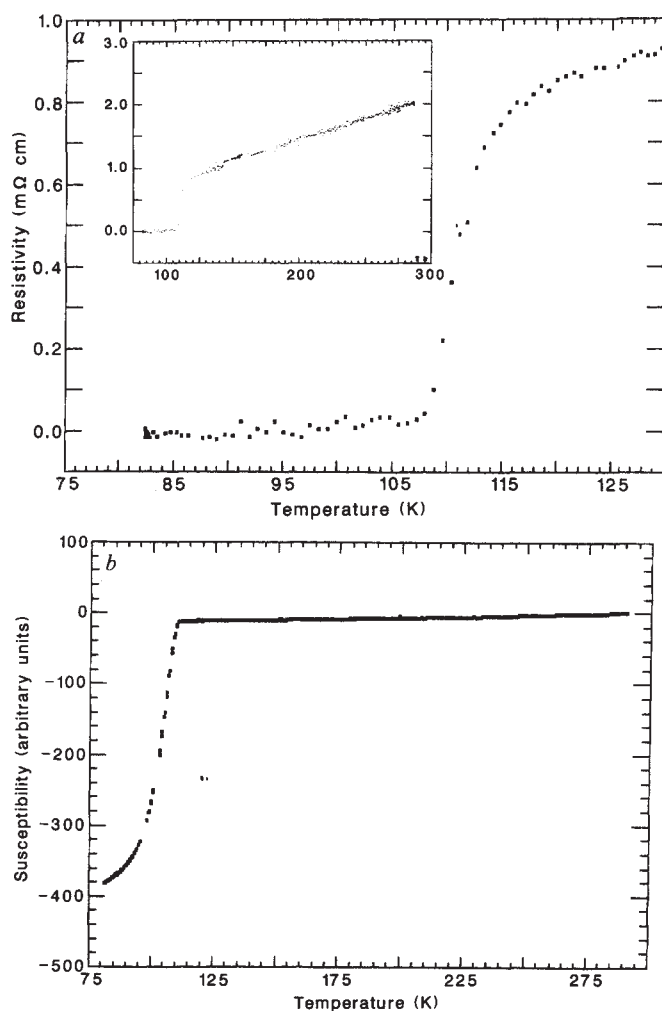


Fig. 1 *a*, Four-point a.c.-resistivity versus temperature. The inset shows the temperature region 75–300 K. *b*, Temperature dependence of the a.c.-susceptibility at 90 Hz.

that this mixture contains a new unknown structure related to the structure of $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$.

The sample was prepared by a solid-state reaction of a mixture with nominal composition $\text{BiSrCaCu}_2\text{O}_x$ from CaCO_3 (99.99%), SrCO_3 (99.9%), Bi_2O_3 (99.5%) and CuO (99.99%). The mixture was heated at 850 °C for 24 h. After grinding and pelletization it was sintered at 880 °C for 20 h, followed by further grinding and pelletization. Next, the material was sintered at 880 °C for 63 h, at 890 °C for 20 min and at 880 °C for 9 h, followed by furnace cooling. All heating was done in air. The sintering temperature was found to be extremely important; another position in the furnace gave a lower fraction of material exhibiting T_C of 110 K. Further details are given in ref. 6.

Analytical electron microscopy and electron diffraction were performed on a Philips EM400, equipped with an elemental analysis system of Tracor. The specimens for this electron microscopy study were prepared by putting a few droplets of a suspension of the ground material on a porous carbon-coated film, supported by a silver grid. The electron diffraction spectrum provided well-separated peaks for all elements (including silver). A proper ratio of the k -factors of the elements, was obtained from the compounds Ca_2CuO_3 , Sr_2CuO_3 and Bi_2CuO_4 .

The superconducting transition temperature, T_C , is observed by four-point a.c.-resistivity measurements in a simple liquid-nitrogen Dewar. The resistivity curve for the multi-phase sample is shown in Fig. 1*a*. The absolute value of the resistivity is estimated to be $\sim 2.7 \text{ m}\Omega \text{ cm}$ at room temperature. The resistance ratio $R(300 \text{ K})/R(125 \text{ K})$ is ~ 2.5 , a value similar to that of

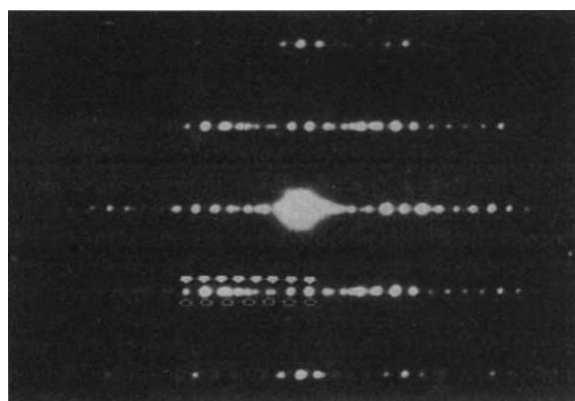


Fig. 2 Electron diffraction pattern of the [100] zone. $\{01l\}$ reflections (all with $l=2n+1$) are given for the 3.7-nm and 3.1-nm structures with full and open white arrows respectively. The reflections from the 3.7-nm phase are much more extended, indicating that this is the majority phase in this selected region.

$\text{YBa}_2\text{Cu}_3\text{O}_7$ polycrystalline sintered material. Figure 1*a* shows a clear metallic behaviour. A linear temperature dependence is observed in the temperature range 300–125 K. A similar linear dependence was observed in $\text{YBa}_2\text{Cu}_3\text{O}_7$ (ref. 7). The deviation towards zero begins at $\sim 125 \text{ K}$, well above T_C . The superconducting transition temperature is as high as 110 K (midpoint). Note that, despite the sharp transition at 110 K, there is still some very small but measurable residual resistance above 95 K.

Figure 1*b* shows the temperature dependence of the real part of the a.c.-susceptibility at 90 Hz. A very sharp transition at 110 K is observed. The diamagnetic signal below T_C is large, suggesting bulk superconductivity. No evidence was observed for an additional superconducting transition at 85 K, often seen in samples given a different heat treatment. The a.c.-susceptibility of the latter samples shows a very sharp T_C at 110 K and a sharp change in the slope of the susceptibility curve at $\sim 85 \text{ K}$. It is noted that the a.c.-susceptibility is proportional to the shielding factor. Thus, in the case of material $T_C = 85 \text{ K}$ surrounded by material $T_C = 110 \text{ K}$, no sign of the $T_C = 85 \text{ K}$ phase would be obtained. More details on electrical resistivity and magnetic susceptibility measurements together with the preparation and characterization of the new superconducting system will be published elsewhere⁶.

Elemental analysis shows $\text{BiSrCaCu}_2\text{O}_x$ to consist of two quaternary phases with the approximate compositions $\text{Bi}_2\text{Sr}_{1.7}\text{Ca}_8\text{Cu}_2\text{O}_x$ and $\text{Bi}_2\text{Sr}_{1.4}\text{Ca}_{1.2}\text{Cu}_3\text{O}_x$, the latter being the minority phase (30%); both normalized to the two bismuth atoms in the formula. A considerable variation in the composition was observed suggesting that at least Ca and Sr can replace each other as can Bi and Cu. Furthermore, a phase containing only calcium was present and there was a very Cu-rich phase (probably an intergrowth of CuO and one of the quaternary compounds).

High-resolution electron microscopy was carried out with a Jeol JEM200CX electron microscope, equipped with a top-entry stage with a $\pm 10^\circ$ double tilt, operating at 200 kV. For this study, small wedge-shaped particles were mounted on a copper slot or small mesh copper grid with silver epoxy glue, in such a way that the thin edge was overhanging the hole. Grinding and mounting of the plate-like material on porous films did not provide particles with the orientation of the plates parallel to the electron beam.

High-resolution electron microscopy showed two types of phases to be present with large c -axis. One of them has diffraction patterns identical to those observed for $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$, showing a tetragonal unit sub-cell of $0.38 \text{ nm} \times 3.1 \text{ nm}$, with a long period modulation in the (110) plane making an angle of

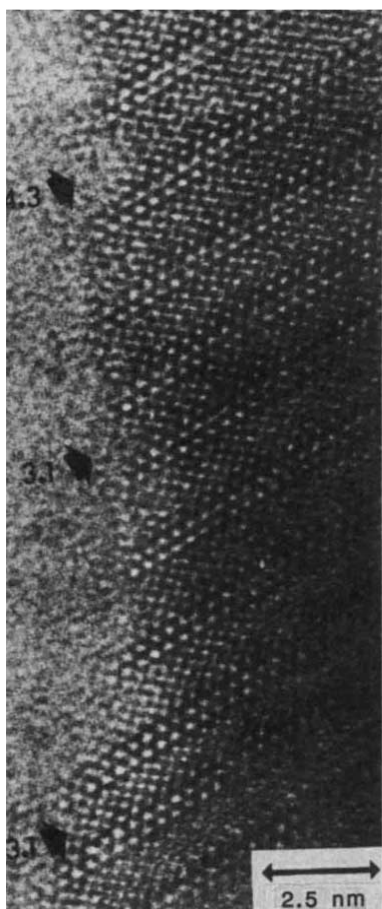


Fig. 3 High-resolution image along [100] of a crystal with a majority of 3.7-nm phase. Intergrowths of the 3.1-nm and 4.3-nm phases, being half a unit cell wide, are indicated. The black dots near the two dominant rows of white dots are the Bi atoms.

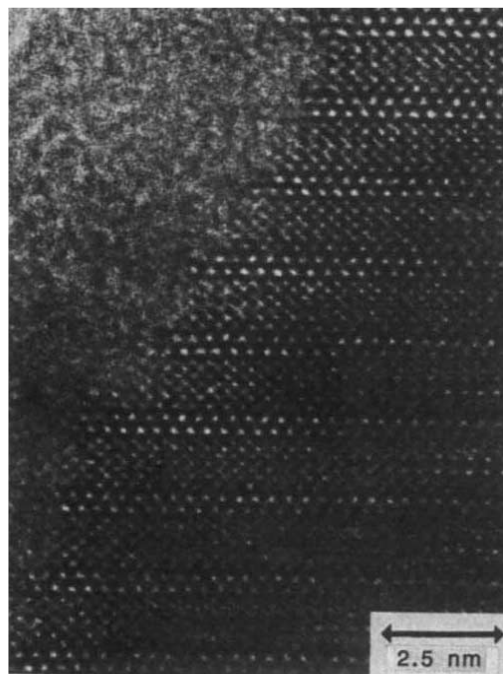


Fig. 4 High-resolution image of the 3.7-nm phase, viewed along [100]. The individual atom columns can be seen. They are schematically given in Fig. 5d.

$\sim 50^\circ$ with the c -axis and with a period of 1.97 nm. High-resolution electron microscopy results on this phase are given in (H. W. Z. *et al.*, manuscript in preparation).

Another phase in the mixture has similar a - and b -axes, but the c -axis is considerably longer, 3.7 nm. This phase was found to be intergrown with the $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$ phase as can be seen in Figs 2 and 3.

High-resolution images of the phase with the longer c -axis are given in Figs 3 and 4. Figure 3 shows the intergrowth of other units in this phase. The intergrowths present can be correlated with a c -axis of 3.1 nm, indicating the $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$ phase, and a c -axis of ~ 4.3 nm. It is evident that the structures of the 3.7-nm and 4.3-nm phase are closely related to the known 3.1-nm phase.

Although individual columns of atoms can be seen in the thin parts of the crystal shown in Fig. 3 they are more evident in Fig. 4. This shows a region of pure phase with a c -axis of 3.7 nm. Individual atom columns along the viewing direction are seen as black dots, which are schematically represented in Fig. 5d. No image calculations have been performed yet. Nevertheless, the structure can be read from the positions of the individual atom columns.

Comparing the pattern of black dots with the pattern of the intergrown $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$ in Fig. 3 it is clear that in the 3.7-nm phase two extra layers of atom columns are inserted normal to the c -axis. The intensity of the dots marking the extra atom columns is not significantly different from the surrounding ones. This implies that they are due to either Cu, Ca or Sr atoms or a combination of them. An important restriction in the com-

parison of the intensities of the black dots is the unusually thick amorphous surface layer, which seems to be inherent to this material.

Two complementary data sets are available to determine more precisely the nature of the inclusion: the increase in the length of the c -axis and the chemical composition. The increase in the c -axis due to the insertion is relatively small, 0.30 nm for each inclusion, and, consequently, 0.60 nm for the c -axis. For an undistorted perovskite block an increase of 0.385 nm is expected, and this kind of insertion can be ruled out, because it would lead to a c -axis of 3.9 nm.

The inclusion of one CuO_2 layer and a Ca layer, similar to the two extra layers of $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$ with respect to $\text{Bi}_2\text{Sr}_2\text{CuO}_{6+\delta}$ will lead to a smaller increase in the length of the c -axis. The composition with this insertion is $\text{Bi}_2\text{Sr}_2\text{Ca}_2\text{Cu}_3\text{O}_{10+\delta}$. The difference in the length of the axes of $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$ and $\text{Bi}_2\text{Sr}_2\text{CuO}_{6+\delta}$ shows that the increase due to the insertion in $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$ is 0.31 nm. Regarding the increase in c -axis, this type of inclusion is possible. Another type of inclusion, two extra CuO_x layers, leading to a composition $\text{Bi}_2\text{Sr}_2\text{CaCu}_4\text{O}_{10+\delta}$, is also in accordance with the high-resolution images and with the increase in the c -axis. To maintain the symmetry of the structure the layer sequence should be $\text{Bi}_2\text{O}_2\text{-SrO-CuO}_x\text{-CuO}_x\text{-Ca-CuO}_x\text{-CuO}_x\text{-SrO-Bi}_2\text{O}_2$. CuO double layers do exist, for instance in SrCuO_2 (ref. 8) and in defects in $\text{YBa}_2\text{Cu}_3\text{O}_7$ ⁹. Elemental analysis suggests that the composition is $\text{Bi}_2\text{Sr}_{1.4}\text{Ca}_{1.2}\text{Cu}_3\text{O}_x$. Taking into account the possible errors in the determination and the possible variation of the composition and the fact that intergrowth with $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$ is often observed, these measurements are in favour of the insertion of two CuO layers but the insertion of a CuO_2 and Ca layer cannot be ruled out. If the structure of the CuO double layer is similar to the ones in SrCuO_2 and $\text{YBa}_2\text{Cu}_3\text{O}_7$ the image along [010] should be different. No data on this direction are available yet.

The stacking defect with a c -axis of 4.3 nm, shown in Fig. 3, is compatible with the inclusion of two more layers. Because of the similarities, this inclusion is most likely to be the same as in the 3.7-nm structure.

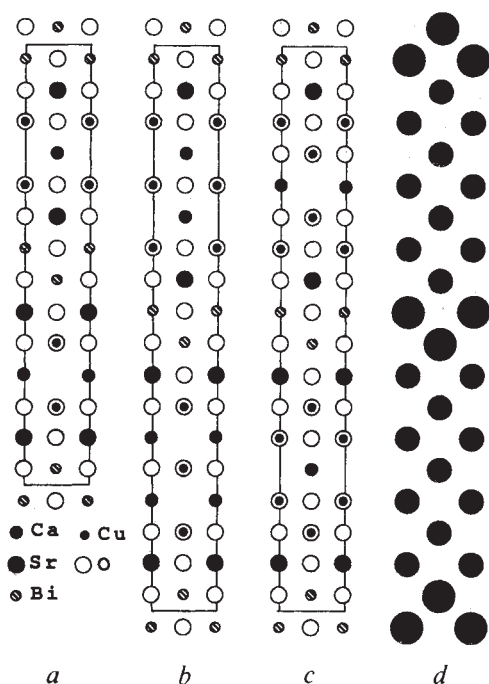


Fig. 5 Schematic representations of the structures of *a*, $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$; *b*, $\text{Bi}_2\text{Sr}_2\text{Ca}_2\text{Cu}_3\text{O}_{10+\delta}$; and *c*, $\text{Bi}_2\text{Sr}_2\text{CaCu}_4\text{O}_{10+\delta}$. *d*, The black dot pattern observed in Fig. 4. The unit cells are indicated by a full drawn line.

The main problem with the 110-K superconductor seems to be the difficulty of synthesis. This is in accordance with the long periodic structures of the two structure types, reported here. In general, long periodic structures are more difficult to synthesize. However, the main problem in the synthesis is believed to be related to the temperature region in which these structures are formed, and the synthesis route used so far. Starting with Bi_2O_3 the mixture has to be reacted at $\sim 800^\circ\text{C}$ to prevent Bi_2O_3 from melting. At this temperature $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$ is readily formed. This implies that, once the material is heated at 880°C , $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$ has to react with the remaining Ca compounds and Cu compounds. Such a process will take a very long time (of the order of days) because of the large particle sizes and because $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$ is a stable compound.

It is concluded that in a mixture with nominal composition $\text{BiSrCaCu}_2\text{O}_x$ exhibiting nearly zero resistance at 107 K, two quaternary compounds are found with the approximate compositions $\text{Bi}_2\text{Sr}_{1.7}\text{Ca}_8\text{Cu}_2\text{O}_x$ and $\text{Bi}_2\text{Sr}_{1.4}\text{Ca}_{1.2}\text{Cu}_3\text{O}_x$, the latter being the minority phase. High-resolution electron microscopy shows the existence of a phase with a *c*-axis of 3.7 nm next to a phase with a *c*-axis of 3.1 nm; the latter, if prepared as a pure phase exhibits a T_C of 85 K. Also an intergrowth occurs pointing to the possible existence of a phase with a *c*-axis of 4.3 nm. The structures of the 3.7- and 4.3-nm phases are strongly related to the structure $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$.

Two indications are obtained that the 3.7-nm phase is responsible for the T_C of 110 K. First, this 3.7-nm phase is only observed in material exhibiting a T_C of 110 K. Second, because the structure is similar to that of $\text{Bi}_2\text{Sr}_2\text{CaCu}_2\text{O}_{8+\delta}$ a high T_C of the 3.7-nm phase is not surprising, the more so because the removal of two perovskite layers as in $\text{Bi}_2\text{Sr}_2\text{CuO}_{6+\delta}$ leads to a drop in T_C to 20 K (ref. 2).

The existence of the 3.7-nm phase shows that in contrast with the structure of $\text{YBa}_2\text{Cu}_3\text{O}_7$, the number of perovskite layers can be changed. This is an important structural variation and it is expected to contribute significantly to the understanding of the principles of high-temperature superconductivity.

Received 26 February; accepted 18 March 1988

1. Dorrian, J. F., Newnham, R. E., Smith, D. K. & Kay, M. I. *Ferroelectrics* **3**, 17-27 (1971).
2. Michel, C. *et al.* *Z. Phys.* **B68**, 421-423 (1987).
3. Maeda, H., Tanaka, Y., Fukutomi, M. & Asano, T. *Jap. J. appl. Phys. Lett.* **27** (in the press).
4. Tarascon, J. M. *et al.* preprint.
5. Zandbergen H. W., Groen W. A., van Tendeloo G., van Landuyt J. & Amelinckx S. *Solid St. Commun.* **66**, 397-403 (1988).
6. Huang, Y. K. *et al.* *J. Cryst. Growth* (submitted).
7. Kadowaki, K., Huang Y. K., van Sprang M. & Menovsky A. A. *Physica* **145B**, 1-4 (1987).
8. Teske C. L. & Mueller-Buschbaum H. Z. *Anorg. Chem.* **379**, 234-241 (1970).
9. Zandbergen, H. W., Gronsky R., Wang K. & Thomas G. *Nature* **331**, 596-598 (1988).

Bulk superconductivity in $\text{Ti}_2\text{CaBa}_2\text{Cu}_2\text{O}_{8+\delta}$ up to 120 K

L. Gao*, Z. J. Huang*, R. L. Meng*, P. H. Hor*, J. Bechtold*, Y. Y. Sun*, C. W. Chu*, Z. Z. Sheng† & A. M. Hermann†

*Department of Physics, Texas Center for Superconductivity and Space Vacuum Epitaxy Center, University of Houston, Houston, Texas 77004, USA

†Department of Physics, University of Arkansas, Fayetteville, Arkansas 72701, USA

Recently, Sheng and Hermann¹ reported the achievement of zero resistance at $>100\text{ K}$ in the system $\text{Ti}-\text{Ca}-\text{Ba}-\text{Cu}-\text{O}$ (TCBCO). Two superconducting phases were immediately identified² as $\text{Ti}_2\text{CaBa}_2\text{Cu}_2\text{O}_{8+\delta}$ (2122) and $\text{Ti}_2\text{Ca}_2\text{Ba}_2\text{Cu}_3\text{O}_{10+\delta}$ (2223). Here we confirm that zero resistance is achievable above 100 K in 2122-TCBCO synthesized by a fast solid-state reaction, and report a small positive pressure dependence of the transition temperature, as in $\text{Bi}_2\text{CaSr}_2\text{Cu}_2\text{O}_{8+\delta}$ ³.

Samples were prepared at Houston by reacting appropriate amounts of Ti_2O_3 , CaO , BaCO_3 and CuO in two steps. Extreme care was taken, because of the toxicity of Ti and Ti_2O_3 . A $\text{Ba}_2\text{Cu}_3\text{O}_5$ powder was made by heating a mixture of BaCO_3 and CuO powders in air at 920°C for 2 h. The resulting powder was mixed with an appropriate amount of Ti_2O_3 and pressed into pellets of diameter 5 mm and thickness 1.5 mm. The pellets were put into a furnace which had been heated to $880-950^\circ\text{C}$, and were heated for 2-5 min in flowing oxygen. Finally, the pellets were quenched in air to room temperature by taking them from the furnace or cooled to room temperature in the furnace by turning off the power with the furnace door slightly open or closed. X-ray powder diffraction was done with a Rigaku D/MAX IIB automated diffractometer with a Cu source. Bar-shaped samples measuring $4 \times 1.5 \times 1.5\text{ mm}$ were cut from the pellets for electrical and magnetic measurements.

Seven TCBCO samples with nominal composition $\text{Ti}_2\text{Ca}_2\text{Ba}_2\text{Cu}_3\text{O}_{10+\delta}$ were synthesized under different conditions

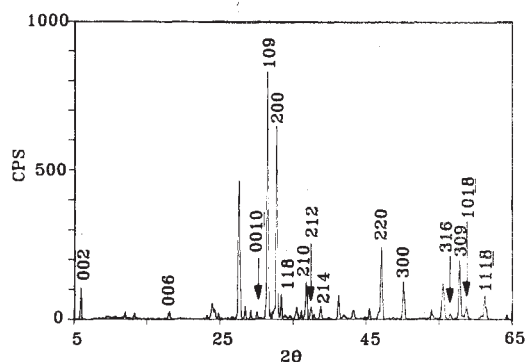


Fig. 1 X-ray powder diffraction pattern of TCBCO sample 3. The peaks arising from the 2122 phase are indexed in a subcell of $5.459 \times 5.459 \times 29.53\text{ Å}$.

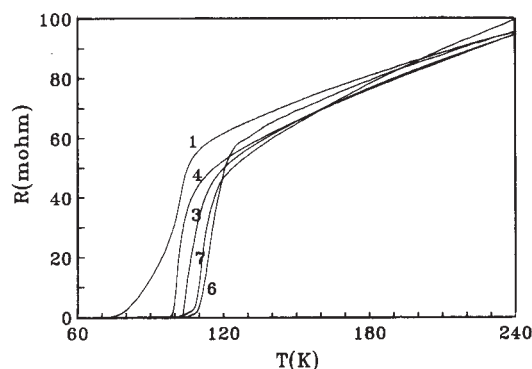


Fig. 2 Resistance versus temperatures for TCBCO samples prepared under various conditions described in the text.

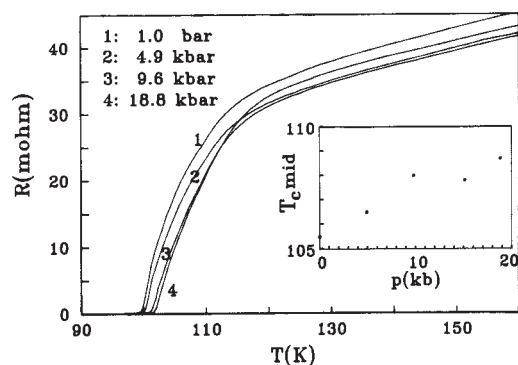


Fig. 3 Resistance versus temperature for TCBCO sample 4 at various pressures. The inset shows the variation of the midpoint of the resistive transition with pressure.

and investigated. The reaction temperatures were 890, 925, 950 and 960 °C for samples 1, 2, 3 and 4–7, respectively. The reaction times were 4, 5, 3 and 2 min for samples 1, 2–5, 6 and 7, respectively. Samples 1–3 were taken from the furnace and quenched to room temperature in air; 4–6 were cooled to room temperature in the furnace with the door slightly open for ~2 h, and sample 7 was cooled to room temperature in the furnace with the door closed for 5 h. Despite the initial stoichiometry of 2223, the X-ray powder diffraction pattern of TCBCO sample 3 is that of the 2122 phase², with a subcell of $5.459 \times 5.459 \times 29.53$ Å, together with an unidentified line at $2\theta = 27.60^\circ$ and small amounts of unreacted oxides, as shown in Fig. 1.

Figure 2 shows the temperature dependence of the resistance for various TCBCO samples. The results show that a high reaction temperature (950–960 °C) and a short reaction time (~3–5 min) are sufficient to obtain zero resistance above 100 K in TCBCO, as previously reported¹. The cooling conditions seem to play only a minor role in determining the character of the resistive transition. The amplitude of the Meissner signal in these samples is about half that in a well prepared $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ (123–YBCO) superconductor, demonstrating that 2122–TCBCO is a bulk superconductor. But the magnetic transition is broadened by a magnetic field as low as 100 Oe, suggesting that weak links are present in the sample.

Figure 3 shows the temperature dependence of resistance at several pressures. The transition temperature (T_c) increases with pressure at a rate of 0.17 K kbar^{-1} , similar to that seen in 2122–BCSCO³ and 123–YBCO⁴ but 5–10 times smaller than that for the 30-K and 60-K oxide superconductors^{5,6}.

The achievement of zero resistance above 100 K in 2122–TCBCO but not in 2122–BCSCO, in spite of the similar Meissner signals observed strongly suggests that there exists a subtle

difference in the growth processes of the higher phase in the two compounds. We have also found that the superconducting properties of TCBCO are enhanced by synthesis in a pure oxygen atmosphere, but this is not the case for BCSCO. As the highest oxidation states of Tl and Bi are +3 and +5, respectively, this difference in behaviour may indicate the importance of a trivalent element in high- T_c oxides.

Work done at Houston is supported by NSF, NASA, the Texas Center for Superconductivity and the T. L. L. Temple Foundation.

Received 3 March; accepted 21 March 1988.

1. Sheng, Z. Z. & Hermann, A. M. *Nature* **332**, 138–139 (1988).
2. Hazen, R. M. *et al.* *Phys. Rev. Lett.* (submitted).
3. Chu, C. W. *et al.* *Phys. Rev. Lett.* (in the press).
4. Hor, P. H. *et al.* *Phys. Rev. Lett.* **58**, 911–912 (1987).
5. Chu, C. W., Huang, Z. J., Meng, R. L., Gao, L. & Huang, Z. J. *Science* **235**, 567–569 (1987).
6. Chu, C. W., Huang, Z. J., Meng, R. L., Gao, L. & Hor, P. H. *Phys. Rev. B* (submitted).

Estimating Southern Ocean eddy flux of heat and salt from satellite altimetry

T. Keffer & G. Holloway*

School of Oceanography; WB-10, University of Washington, Seattle, Washington 98195, USA

* Institute of Ocean Sciences, P.O. Box 6000, Sidney, British Columbia V8L 4B2, Canada

An outstanding problem in the oceanic and atmospheric sciences is the rate of heat and fresh water transport from the equator to the poles, for it is this transport which powers the Earth's weather and climate system¹. Closing the Southern Hemisphere balances has proven particularly difficult because of the dominance of eddy fluxes across the Antarctic Circumpolar Current (ACC). Their relatively small space, but long time scales, make any *in situ* measurement program prohibitively expensive. Here we use satellite data to estimate the overall eddy activity and a turbulence closure scheme to relate this to eddy flux. The resulting estimated heat flux is sufficient to supply the heat lost to the atmosphere south of the Antarctic Polar Front. Estimated salinity flux is found to be poleward at high latitude—in the wrong direction to compensate for the observed excess of precipitation over evaporation south of the Polar Front.

In an indirect approach, Gordon and Owens² have estimated that a cross-Polar Front heat flux of $0.31 \times 10^{15} \text{ W}$ must be necessary to supply the heat lost to the atmosphere in the Antarctic region south of the Polar Front. Many other investigators have made indirect estimates^{2–8} (see Table 1) but confirmation by direct estimates has proved much more difficult. This is so partly because of logistical reasons—deployment of instruments in the ocean is always an expensive and difficult proposition—but mostly because of dynamic reasons. Eddy fluxes are proving to be the dominant component of the north-south transport in the zonally unbounded Southern Ocean, just as they are in the zonally unbounded atmosphere⁹. The relatively small spatial dimensions of oceanic eddies (15–50 km) are two orders of magnitude smaller than in the atmosphere. There may be literally hundreds across an ocean basin, making the measurement of their accumulative flux a formidable undertaking.

While the eddy flux can be elusive, DeSzoek and Levine⁷ have made an estimate of the heat flux carried by steady (that is, mean) motions alone. This latter flux can be divided into a barotropic ('horizontally' overturning) mode and a baroclinic ('vertically' overturning) mode. DeSzoek and Levine note that across a path of constant vertically integrated temperature, the heat transport by the barotropic mode vanishes, for there will

Table 1 Estimates of cross-Antarctic Polar Front heat flux

Investigator	Value		Method
Newton ³	0.29×10^{15} W	(I)	Subjective balance from satellite radiometry, atmospheric transport, and ocean-atmosphere heat fluxes.
Gordon and Owens ²	0.31×10^{15} W	(I)	Global abyssal ocean heat balance.
Bryden ⁴	0.54×10^{15} W	(D)	Direct measurements in Drake Passage of eddy heat flux only.
Trenberth ⁵	1.0×10^{15} W	(I)	Satellite radiometry.
Hastenrath ⁶	0.13×10^{15} W	(I)	Integration of air-sea exchange from the North Pole southwards.
DeSzoeke and Levine ⁷	-0.15×10^{15} W	(D)	Hydrographic estimates of barotropic (see text), baroclinic and Ekman fluxes alone.
Gordon ⁸	0.54×10^{15} W	(I)	Integration of air-sea exchange from the South Pole northward to 60° S.

I, indirect estimate; D, direct estimate.

be no along-path temperature variability to correlate with the vertically uniform velocity. By restricting their calculation to such a path they are able to avoid completely the problem of estimating a reference velocity in order to calculate the barotropic velocity field. They are then free to use the dynamical method to calculate the baroclinic heat flux. They did this along the path of 1.3 °C vertically averaged temperature, roughly corresponding to the position of the Antarctic Polar Front, and obtained an interesting and surprising result: zero. To this they add an estimate of the ageostrophic Ekman flux of 0.15×10^{15} W to the north, giving a total poleward cross-ACC flux of -0.15×10^{15} W. They hypothesize a missing 0.45×10^{15} W poleward eddy flux to maintain the heat balance suggested by Gordon and Owens². It should be noted that their calculations did not include the transport of heat by deep boundary currents, an alternative to eddy fluxes.

Multi-year deployments of current meter moorings within the Drake Passage during the International Southern Ocean Studies program allowed Bryden⁴ to calculate the eddy heat flux directly as $c_p \bar{v'T'}$, where c_p is the specific heat of water, v' and T' are the fluctuating (eddy) components of northward velocity and temperature, respectively, and the overbar indicates a time average. He found a statistically significant temporal correlation between velocity and temperature and, from this, estimated a local poleward heat flux of 0.67 W cm^{-2} . If this heat flux were acting uniformly over an ocean depth of 4 km and a 20,000 km section (as suggested by Bryden), then this would correspond to a total poleward heat flux of 0.54×10^{15} W, sufficient to supply the observed Antarctic heat losses. This technique is only locally accurate and there are questions about just how representative the Drake Passage is of the Southern Ocean as a whole. But the task of ringing the Antarctic with current meter moorings to obtain an accurate total flux is daunting.

We explore a very different approach, using satellite measurements of sea surface variability to estimate horizontal eddy diffusivity which can then be multiplied by mean temperature and salinity gradients to estimate eddy fluxes. There is a great deal of uncertainty in the approach. At present, the important question is whether the results 'make sense' and, if they do, whether any further effort is warranted to refine the precision of the method.

Our crucial assumption is that under conditions of statistically homogeneous, barotropic, β -plane turbulence, the eddy diffusivity tensor κ can be estimated as $\kappa = C\psi\tau$, where ψ is the r.m.s. geostrophic stream function, C is a constant or proportionality (~ 0.4), and τ is an $O(1)$ tensor describing the anisotropy induced by Rossby wave propagation^{10,11}. The meridional component of τ is given by $(1 + (\mu\beta)^2)^{-1}$, where $\beta = \beta u/\zeta^2$. Here, β is the meridional derivative of the Coriolis parameter f , and u and ζ are the r.m.s. eddy flow speed and vorticity, respectively. The factor μ depends on the details of the form of the eddy spectrum. Given an estimate of κ , the meridional eddy heat flux

per cross-sectional area can be estimated as $F_T = -c_p \kappa_y (\partial \bar{T}/\partial y)$ where $\bar{T}$ is depth-averaged temperature. The eddy flux of salt can be estimated in a similar manner.

Values of μ and β will vary with geographic location and with the seasons in ways that are not well determined. We emphasize that these values, and also C , are not known with confidence. This could invite 'tuning' of the adjustable parameters, a hazard we seek to avoid. For the purposes of demonstration and consistency only, we assign exactly the values used previously by Holloway¹⁰: $\mu = 3$ and $\beta = \frac{1}{3}$.

The r.m.s. eddy stream function ψ is derived from 24 days of repeat tract altimetry of sea surface height (after removing tides and correcting for atmospheric, sea state and orbital variations), as measured by SEASAT, a short-lived (3.5 month) satellite launched in 1978, by using the relationship $\psi = gh/f$, where h is the r.m.s. sea surface height and g is gravity. Because of the short duration of the tracks, only part of the total eddy variability was measured. By comparison with spectra from long-term moorings and with previous GEOS 3 altimetry, it has been estimated¹² that the fraction was about $\frac{1}{3}$, requiring that the measured estimates of h be multiplied by three. On the other hand, the satellite sees only surface variability which typically decays with depth. We assume that this variance at depth is one half the surface variance. The net correction to h (and, therefore, ψ) is a factor of $3 \times \sqrt{\frac{1}{2}} \approx 2$.

Again, we emphasize the tentative character of these assignments. Present (for example, GEOSAT) and future satellites will estimate the surface variability more completely. Preliminary examination of GEOSAT results indicate that the fraction of long-term variability seen by SEASAT was more like $\frac{1}{2}$ in the ACC region and 1 in the mid-ocean, instead of the $\frac{1}{3}$ used here and in previous works¹². Observations during the Mid-Ocean Dynamics Experiment indicate that the functional dependence of κ on ψ may be about right, but that ψ variance may be more strongly surface intensified than we have assumed¹³. Moreover, we seek to calculate eddy transport of heat, a dynamically active quantity, while the theory yielding $\kappa = C\psi\tau$ assumes dynamic passivity. Heat anomalies are actively expressed in the surface elevation field. Geostrophic dynamics tend to steer ocean currents along isotherms, hence limiting the effectiveness of eddy heat transport. As in the previous work¹⁰, we have sought to discount this inefficiency by calculating eddy flux as the product of depth-averaged diffusivity times depth-averaged property gradient, offsetting some of the uncertainty in depth distribution of ψ variance. Finally, assumptions of geostrophy break down at low latitudes and estimates of sea-surface variability are invalid over sea ice at high latitudes.

The data source to calculate $\partial \bar{T}/\partial y$ (and $\partial \bar{S}/\partial y$) was the climatological atlas of Levitus¹⁴. We averaged the data from the surface to 2,000 m. Where the bottom depth is less than 2,000 m (a small fraction of the total area), the location was ignored. The resulting two-dimensional grid was finite-differenced in y

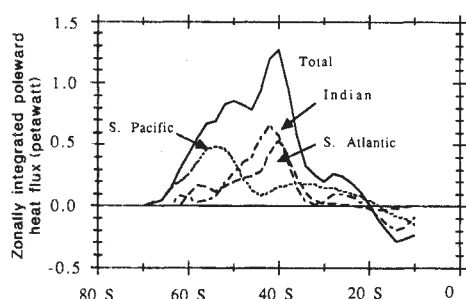


Fig. 1 Zonally integrated poleward eddy heat flux in petawatts (10^{15} W).

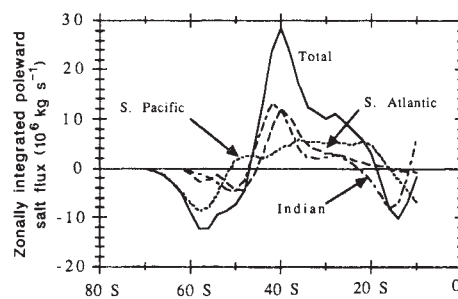


Fig. 2 Zonally integrated poleward eddy salt flux in units of 10^6 kg s^{-1} .

and multiplied by the local estimate of κ_y to give local estimates of eddy flux. These were then zonally integrated to give the total meridional heat flux, latitude by latitude. This procedure was repeated for salinity. The results are shown in Figs 1 and 2.

We estimate a cross-ACC eddy flux of 0.7×10^{15} W at 53° S, the mean latitude of the Polar Front. If this heat flux is distributed uniformly over ocean 4,000 m deep and 20,000 km across, then this corresponds to an average flux of 0.87 W cm^{-2} .

The heat and salt fluxes of the South Pacific are shifted southwards relative to those of the other oceans because of the relatively southern position of the Polar Front in the Pacific. The position and size of the peaks seen in the Indian and South Atlantic Oceans are due to a combination of the relatively northern position of the Polar Front ($\sim 53^\circ$ S) and (especially) the very active Agulhas and Falkland Confluence regions. In all oceans the heat flux is towards the equator at low latitudes (10 – 20° S).

At subtropical latitudes (~ 10 – 20° S) the salt flux is generally away from the gyres. From compilations of net evaporation minus precipitation¹⁵, we estimate a net flux of $\sim 6.6 \times 10^8$ kg s^{-1} of freshwater from ocean to atmosphere within these latitude bands. At an average salinity of 35 p.p.t., this represents an equivalent salt flux of 23×10^6 kg s^{-1} into the ocean. The eddies appear capable of disposing this salt southward, across the southern boundaries of the subtropical gyres, where it is diluted by a net precipitation over evaporation.

Gordon and Taylor¹⁶ estimate an input of 0.32×10^6 m³ s^{-1} of freshwater to the ocean south of the Polar Front, due to runoff and an excess of precipitation. To dilute this input to 35 p.p.t. will require a poleward flux of 11×10^6 kg s^{-1} of salt across the Polar Front. We find, however, an eddy flux that is uniformly equatorward at these latitudes. This direction of transport occurs because the horizontal salinity gradient is generally poleward at all depths below 200 m, due to the southward infusion of salty North Atlantic Deep Water and Circumpolar Deep Water along tilting isopycnals and into the Antarctic zone. Thus, the ocean is in the interesting configuration of having poleward horizontal salinity gradients, but equatorward along-isopycnal gradients.

Ekman transport can also contribute to the flux of salt. Above 200 m the salt gradient is equatorward, as is the Ekman transport. This implies an equivalent poleward salt transport. From our hydrographic data we estimate an average salinity of 34 p.p.t. in the upper 200 m, 34.5 p.p.t. for the upper 2,000 m, both values taken at the Polar Front. Assuming an Ekman volume transport of 28×10^6 m³ s^{-1} equatorwards⁸ at 34 p.p.t. with a return flow at 35.5 p.p.t., this represents an equivalent poleward salt flux of 14×10^6 kg s^{-1} , about right to balance the observed excess of precipitation, but not the estimated eddy fluxes. We conclude that either along-isopycnal or baroclinic fluxes of salt must be important.

The analysis we have presented here has promising features. At least in gross, a major imbalance between the previously calculated mean ocean heat transport and the heat transport required by regional budget has been closed by an estimation

of the eddy transport. Salt transports are plausible over most latitudes but are less-successfully accounted for by the eddy fluxes at high latitudes. Major uncertainties in the method of eddy transport calculation, discussed in the preceding text; pose a challenge to future research. Present and future satellite missions will enormously expand and refine the data set from which such calculations can be made. We see the results presented here, however tentative, as an encouragement to continue with this approach.

We thank R. Cheney for supplying SEASAT altimetric data. Research was supported in part by the Office of Naval Research and by the NSF.

Received 28 September 1987; accepted 4 February 1988.

1. Charnock, H. Q. *Jl. met. Soc.* **113**, 3–18 (1987).
2. Gordon, A. L. & Owens, W. B. *Rev. Geophys.* **25**, 227–233 (1987).
3. Newton, C. W. *Meteorology of the Southern Hemisphere, Meteorol. Monogr.* **13**, 215–246 (American Meteorological Society, Boston, 1972).
4. Bryden, H. L. *J. mar. Res.* **37**, 1–22 (1979).
5. Trenberth, K. E. *Dyn. Atmos. Oceans* **4**, 57–64 (1979).
6. Hastenrath, S., *J. phys. Oceanogr.*, **10**, 159–170 (1980).
7. DeSzeke, R. A. & Levine, M. D. *Deep-Sea Res.* **28**, 1057–1085 (1981).
8. Gordon, A. L., *J. geophys. Res.* **86**, 4193–4197 (1981).
9. Palmen, E. & Newton, C. W. *Atmospheric Circulation Systems* (Academic, New York, 1969).
10. Holloway, G. *Nature* **323**, 243–244 (1986).
11. Holloway, G. & Kristmannsson, S. *J. Fluid Mech.* **141**, 27–50 (1984).
12. Fu, L.-L. *Rev. Geophys. Space Phys.* **21**, 1657–1666 (1983).
13. Freeland, H. *Nature* **326**, 524 (1987).
14. Levitus, S. *Climatological Atlas of the World Ocean* NOAA Prof. Paper 13, 1973 pp (U.S. Govt. Printing Office, Washington, 1982).
15. Newell, R. E., Vincent, D. G., Dopplick, T. G., Ferruzza, D. & Kidson, J. W. in *The Global Circulation of the Atmosphere* (ed. Corby, G. A.) 42–90 (Royal Meteorological Society, London, 1969).
16. Gordon, A. L. & Taylor, H. W. *Numerical Models of Ocean Circulation* (National Academy of Science, Washington, D.C., 1975).

Field evidence for hydraulic jumps in subaqueous sediment gravity flows

Frank H. Weirich*

Department of Geography, University of California, Los Angeles, California 90024, USA

Despite the significance of subaerial and subaqueous sediment-laden gravity flows as geological and geomorphic agents^{1–5}, our understanding of such transport and depositional processes is limited. In particular, there is little known about the mechanisms involved in the conversion of higher-density subaqueous debris flows into lower-density turbidity currents. To address this issue, a detailed field study of this conversion process was undertaken on flows and their deposits within a reservoir. The results presented here provide field evidence indicating the occurrence of hydraulic jumps as a conversion mechanism in such flows, and the deposition

* Present address: Department of Geography and Department of Civil & Environmental Engineering, and the Iowa Hydraulics Institute, University of Iowa, Iowa City, Iowa, USA.

of large quantities of sediment in the immediate down-jump area. The findings are significant both in terms of basic geological processes as well as in applications to problems of reservoir management in areas with high sediment inputs.

The need to determine the mechanisms by which subaqueous debris flows convert into less dense turbidity currents is important, having direct bearing upon the overall flow dynamics of sediment-laden gravity flows. Consequently, there has been considerable debate as to the conversion processes. Some have suggested that mixing and dilution occurs at the front or head of such flows^{6,7}, and others have argued for entrainment at the interface between the body of the flow and the overlying fluid⁸. A third mechanism of a hydraulic jump has been suggested⁹⁻¹¹, which permits the rapid conversion from a slump to a turbidity current. More recently, Fisher¹² has suggested that all of these conversion mechanisms are possible: body, gravity, surface and fluidization transformations. For the most part the discussion has remained largely theoretical and laboratory-based with only limited stratigraphic and process data available.

Beginning in 1984, I undertook a field study to improve our understanding of both debris flows and turbidity currents, with particular emphasis on the conversion process from one to another. The study was carried out in the San Dimas Experimental Forest located in the San Gabriel Mountains ~80 km east of Los Angeles (Fig. 1). The Experimental Forest is situated in an area noted for extreme gradients (mean slope of 68%)¹³ and often intense rainfall¹⁴, resulting in high sediment yields¹⁵. To increase the likelihood of major debris flows, two watersheds were burned to remove the vegetation cover. The combination of high slopes, intense rainfall and enhanced sediment yield (up to 35-fold increase as a result of fire¹⁶) offered the possibility of producing large-scale debris flows which would travel downstream from the burned basins to San Dimas reservoir (Fig. 1).

At the base of both burned watersheds, large concrete flumes were instrumented to provide continuous measurements of flow discharge. In addition, I designed a new type of density sensor to monitor continuously the concentration of sediment within the flows. The sensors employ a series of solid-state, piezoelectric, differential pressure detectors mounted in the flow to monitor the density of the fluid as it moves through the calibrated flumes. With simultaneous water-level and temperature data, and referencing of pressure levels to the atmosphere, it is possible to determine the component of the fluid density due to the presence of sediment. The data were recorded on a high-speed logging system located near the flume at each site. The instrumentation systems, operating continuously from November 1984 until May 1985, provided a detailed record of the nature and character of debris flows moving out of the burned watersheds. In addition, the draining of San Dimas reservoir in the summer of 1984 enabled a study of the bottom topography before the winter storms.

A number of debris flows occurred following rain events over the period of monitoring. A storm on 19 December 1984 is of particular interest. The peak intensity of 5 mm in 5 min at 16:15 triggered debris flows in both of the burned basins. The front of the debris flow in basin 2 reached the flume at the mouth of the watershed at 16:30, whereas the front of the debris flow in basin 1 reached the flume in that watershed some 3 min later. Both events lasted ~15 min.

In addition to the instrumentation data available, I was fortunate to observe both events. From the outlet of the burned basins, both debris flows followed the main channel to the reservoir (Fig. 1). Observation of the subaerial flows indicated that they consisted of: (1) a head containing a large quantity of organic material, including ash, leaf litter and burnt twigs and branches; (2) a body of material moving as a rather fluid debris flow; and (3) a tailing-off portion of sediment-laden fluid flow.

Data from the sensors installed in both flumes indicated a mean sediment load of ~60% by weight with a total volume of

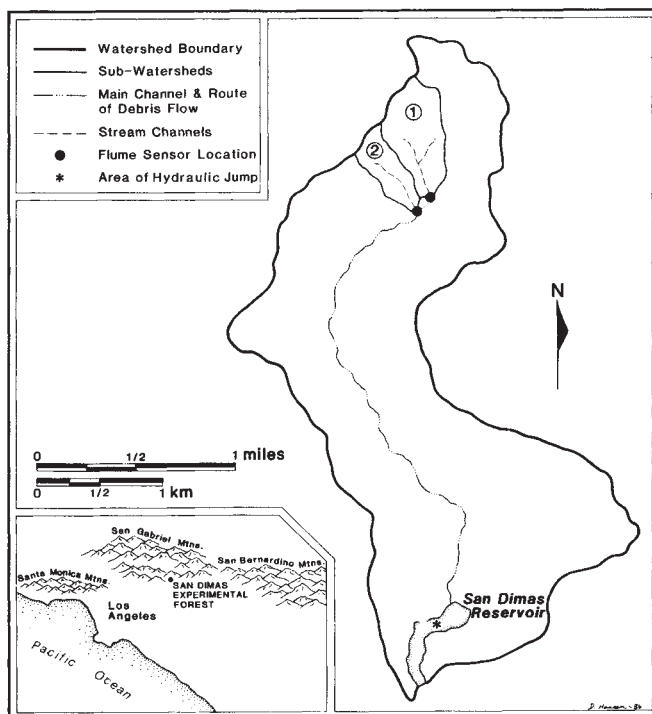


Fig. 1 Location and components of the San Dimas Experimental Forest involved in the study of the subaqueous hydraulic jump. Debris flows passed through instrumented flumes at the base of the two watersheds and then moved down the main channel and out onto the reservoir floor.

sediment of $6.0 \times 10^3 \text{ m}^3$. Surveys of the subsequent deposits indicated that ~50% of the material was deposited in the channel upstream of the reservoir, while the remainder moved out into the reservoir.

Following the 19 December event, no major storms occurred which resulted in debris flows. Thus, evidence of these flows remained intact following the December storm. The reservoir was then lowered in the summer of 1985 exposing the floor and deposits. The presence of fire-related ash and burnt organic material enabled the recognition of the subaqueous deposits from the debris flow events. Moreover, the upper limit of the fire-related materials in the walls of a pre-existing subaqueous channel provided a distinct and readily traceable indication of the actual depth of the flows. Mapping and analysis of the deposits indicated that both flows followed the pre-existing channels in the reservoir floor. The flows moved down this channel for several hundred metres and then fanned out to cover nearly the entire width of the reservoir bottom (Fig. 2a).

The thickness and character of the deposits varied substantially and changed abruptly. For nearly 160 m from the delta front, the subaqueous debris flow deposits maintain a relatively constant thickness, suggesting little change in flow conditions. This supports the findings of laboratory studies⁷, which suggest that subaqueous debris flows are capable of travelling substantial distances without thickening of the flow or entrainment. The abrupt thickening of the deposits, from an average of 0.3 to 1.0 m in the area downstream from the bend in the channel, is of particular note (Fig. 2a and b). Such a rapid change in deposit thickness is indicative of a significant change in the subaqueous flows themselves. A marked increase in deposit thickness is the type of change to be expected after a rapid flow transformation produced by a hydraulic jump, rather than from some gradual dilution and entrainment process.

The actual depth of the underflows, confirmed by the upper limit of the ash deposits on the channel walls, follows a similar pattern. The depth of flow, stable for a distance of nearly 160 m,

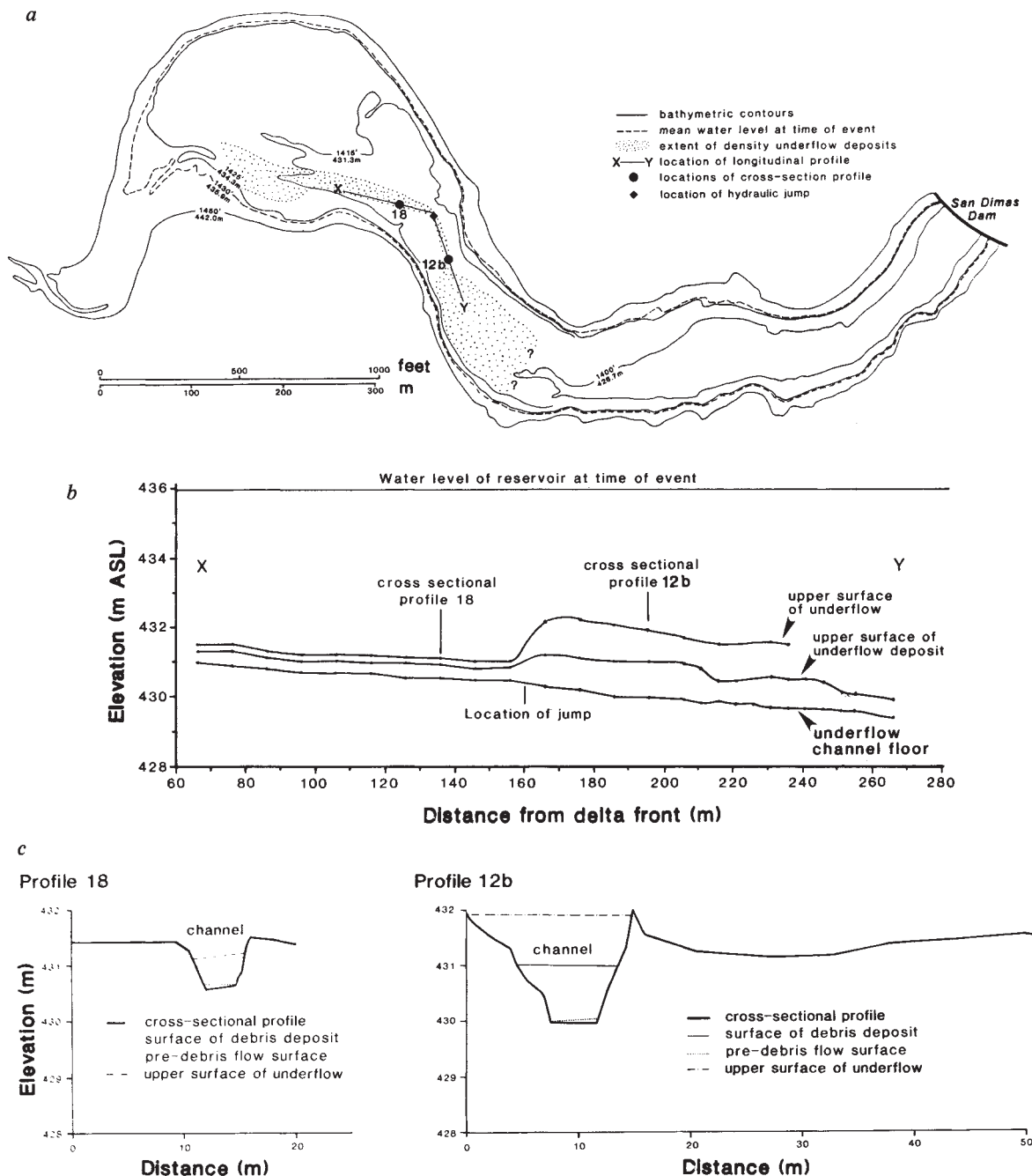


Fig. 2 *a*, A plan view of the San Dimas reservoir indicating the route of the flows and the main deposition zone located well down reservoir. *b*, The longitudinal profile of flow depth, deposit thickness and the floor of the underflow channel indicate the down-reservoir changes in the flow. *c*, The cross-sectional profiles upstream and downstream from the subaqueous hydraulic jump clearly reflect the change in sediment and flow conditions produced by the jump.

increases rapidly from an average depth of 0.5 to nearly 2.0 m in a horizontal distance of less than 10 m (Fig. 2*b*). Moreover, this change in flow depth coincides with the change in deposit thickness. Such sudden and simultaneous changes in both flow depth and deposit thickness, strongly supports the interpretation that a hydraulic jump occurred at this point with a rapid alteration of the flow from supercritical to subcritical conditions.

The equations for subaerial hydraulic jumps are well established¹⁷, but our knowledge of subaqueous processes is quite limited. Applying the equations of Yih and Guda¹⁸ for subaqueous jumps yields a pre-jump Froude value of 3.2, and using the desimetric Froude equation¹¹ indicates a pre-jump velocity of 3.9 m s^{-1} . The post-jump Froude values were subcritical with velocities approaching 1.0 m s^{-1} . The calculated

pre-jump velocity is also consistent with observations made on the velocity of the subaerial debris flows. Moreover, it appears that both subaqueous debris flows underwent a hydraulic jump at about the same point within the pre-existing channel (Fig. 2*b*).

The occurrence of hydraulic jumps is further supported by the stratigraphic evidence within the deposits. The pre-jump deposits upstream from the bend in the channel consist of classic debris flow deposits with no structure and little if any sorting. Conversely, in the post-jump area downstream from the bend in the channel, it was possible to differentiate the two separate debris flows. The deposits in this area consisted of two distinct charred organic layers, each overlaid by deposits from the body of the flows. Within the deposits of both flows, I also found extensive areas of climbing ripples and bedding, indicative of

lower flow regime conditions. This is consistent with the velocities calculated for the post-jump flow. Immediately downstream of the jump area there was a relatively rapid outfall of material such that 70% of the total sediment load carried by the flow was deposited within 200 m of the jump.

The presence of the hydraulic jump and the rapid outfall of material in the down-jump environment is of significance from both a theoretical and applied perspective. From a theoretical viewpoint these findings provide new insight into basic turbidity current generation processes by demonstrating the ability of debris flows to effectively convert to turbidity currents by undergoing a subaqueous hydraulic jump.

The presence of the subaqueous jump, and the related concentration of the deposits in the immediate post-jump environment, also have implications for applied situations. Knowledge of the conditions which are conducive to generating a subaqueous hydraulic jump could assist in the prediction of sedimentation patterns in reservoirs. Moreover, it may be possible to control the location of such jumps and the resulting deposition of sediment by altering the subaqueous bed and channel characteristics. Thus, these findings could prove useful not only in understanding and predicting deposition patterns, but also in enabling reservoir engineers to control the location of deposition.

This study was supported by UCLA Senate Grants, the Los Angeles County Flood Control District and the Los Angeles

County Fire Department. The watershed data were assembled as part of an ongoing study of watershed response to fire with Dr Phillip Riggan of the Riverside Fire Laboratory of the USDA. I also gratefully acknowledge the assistance of Chuck Culver of the United States Forest Service and Tom Ryan, Linda O'Hirok and Kevin Mulligan.

Received 27 July 1987; accepted 21 January 1988.

1. Heezen, B. C. & Ewing, M. *Am. J. Sci.* **250**, 849-873 (1952).
2. Middleton, G. V. & Hampton, M. A. in *Marine Sediment Transport and Environmental Management* (eds Stanley, D. J. & Swift, D. J. P.) 197-218 (Wiley, New York, 1976).
3. Takahashi, T. *Am. Rev. Fluid Mech.* **13**, 57-77 (1981).
4. McCutcheon, S. C. & Bradley, J. B. in *Water for Resources Development* (ed. Schreiber, D. L.) 43-47 (American Society of Civil Engineers, 1984).
5. Pierson, T. C. & Scott, K. M. *Wat. Resour. Res.* **21**, 1511-1524 (1985).
6. Allen, J. R. L. *J. sedim. Petrol.* **41**, 97-113 (1971).
7. Hampton, M. A. *J. sedim. Petrol.* **42**, 775-793 (1972).
8. Morgenstern, N. R. in *Marine Geotechnique* (ed. Richards, A. F.) 189-220 (University of Illinois Press, Urbana, 1967).
9. Menard, H. W. *Marine Geology of the Pacific* (McGraw-Hill, New York, 1964).
10. Van Andel, T. H. & Komar, P. D. *Geol. Soc. Am. Bull.* **80**, 1163-1190 (1969).
11. Komar, P. D. *Geol. Soc. Am. Bull.* **82**, 1477-1487 (1971).
12. Fisher, R. V. *Geol. Mag.* **11**, 273-274 (1983).
13. Hill, L. W. *The San Dimas Experimental Forest* (USDA Forest Service, Berkeley, California, 1963).
14. Sinclair, J. D. *Trans. Am. geophys. Un.* **35**, 264-268 (1954).
15. Scott, K. M. & Williams, R. P. *U.S. geol. Surv. Prof. Pap.* 1030 (1978).
16. Rowe, P. B., Countryman, C. M. & Storey, H. C. *Hydrologic Analysis Used to Determine Effects of Fire on Peak Discharge and Erosion Rates in Southern California Watersheds* (USDA Forest Service, Berkeley, California, 1954).
17. Chow, V. T. *Open Channel Hydraulics* (McGraw-Hill, New York, 1959).
18. Yih, C. S. & Guha, C. R. *Tellus* **7**, 358-366 (1955).

Marine dispersal and deposition of Yellow River silts by gravity-driven underflows

L. D. Wright*, W. J. Wiseman†, B. D. Bornhold‡, D. B. Prior†, J. N. Suhayda†, G. H. Keller§, Z.-S. Yang|| & Y. B. Fan||

* Virginia Institute of Marine Science, College of William and Mary, Gloucester Point, Virginia 23062, USA

† Louisiana State University, Baton Rouge, Louisiana 70803, USA

‡ Geological Survey of Canada, Sidney, British Columbia V8L 4B2, Canada

§ Oregon State University, Corvallis, Oregon 97331, USA

|| Shandong College of Oceanography, Qingdao, China

Field measurements of the processes of dispersion of concentrated suspended silts over the active delta front of the Yellow River (China) show that cross-isobath sediment dispersal into the shallow Bohai Gulf is dominated by gravity-driven underflows. These underflows descend the rapidly prograding delta front as hyperpycnal plumes 1-4 m thick. The widespread, dilute plumes observed during the summer flood season deposited most of their sediment and experienced extinction between the 5-m and 10-m depth contours. High-frequency internal waves on the upper interface of the underflows may have contributed to the rapid deceleration. Muds resuspended by an intense autumn storm caused denser, but short-lived, plumes which extended downslope to depths of over 13 m.

Gravity-driven turbidity currents or turbulent suspension currents have been widely studied¹⁻³. The possible importance of these mechanisms to the dispersal of sediments in shallow water has only recently been seriously considered^{4,5}. Most models for the dispersion of river-borne sediments into the sea have mainly emphasized transport of fines in hypopycnal (buoyant) plumes⁶. But we have discussed the probable importance of a hyperpycnal dispersal mode off the mouth of the silt-charged Yellow River (Huanghe; People's Republic of China)^{5,7} after an initial reconnaissance cruise in the Western Bohai Gulf over the active subaqueous delta of the Yellow River (latitude 37°35' N-

37°50' N; longitude 119°10' E-119°25' E) in May and June 1985. Reports have indicated that the suspended load concentrations issuing from the Yellow River into the shallow (mean depth, 30 m) Bohai average 25 kg m⁻³, and attain flood stage maxima of 220 kg m⁻³ (ref. 8). The suspended solids are mainly loess silts and very fine sands with settling velocities in the 0.5-2.5 × 10⁻³ m s⁻¹ range^{9,10}.

After confirmation that hyperpycnal underflows existed⁵, subsequent cruises in July-August 1986 and September-October 1987 aimed at documenting the areal extent, vertical thickness, bulk densities, downslope velocities and velocity gradients of the underflows and the relationship of underflow behaviour to tidal and storm forcings. Depth-averaged suspended sediment concentrations in the Huanghe near the mouth were about 28 kg m⁻³ during the 1986 cruise; this was below normal for July and August. During the 1987 cruise, discharge was negligible, but two severe storms occurred. We used two ships to make simultaneous observations at the top (depth, 4-5 m) and bottom (depth, 9-11 m) of the steepest (0.6°) slope segment of the active delta front. Time-series stations were occupied for 26 h, during which conductivity, temperature, suspended sediment concentration, and current speed and direction were measured. At the shallow stations the vertical and temporal sampling intervals were 1 m and 1 h; at the deeper stations the respective intervals were 2 m and 2 h. Three pairs of simultaneous stations were occupied during the 1986 cruise. One pair was occupied in 1987. Other individual time-series stations provided extensive spatial coverage.

Strong, reversing isobath-parallel tidal currents flowing along-slope were observed at all time-series stations in 1986 and 1987. These flows often had speeds of over 1.5 m s⁻¹ near the surface and 0.80 m s⁻¹ at 1 m above the bed. At the shallow stations they contributed to resuspension of recently deposited silts and promoted early intense mixing of fresh and salt water. Cross-isobath flow components observed in 1986 had downslope speeds of 0.05-0.30 m s⁻¹ near the bottom. Suspended sediment concentrations in the lower 2 m of the water column at the shallow stations normally exceeded 1 kg m⁻³ and attained maxima of over 10 kg m⁻³. Near-surface concentrations at the 1986 shallow stations varied from <0.10 kg m⁻³ to >1 kg m⁻³, depending on the position of a sharp plunge-point front⁵ which

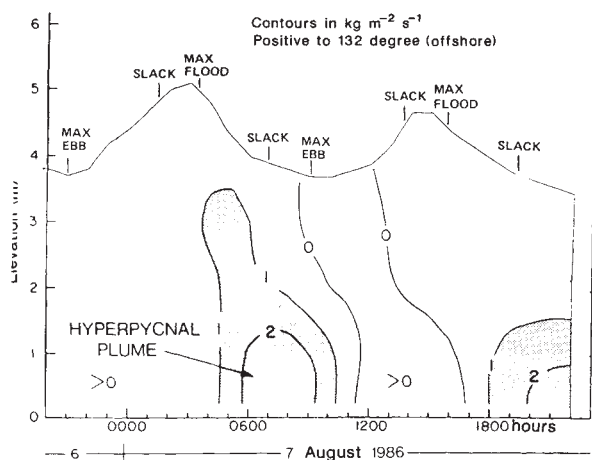


Fig. 1 Tidal cycle (26 h) variations in depth- and time-dependent downslope sediment fluxes (positive to 132°) at shallow Station E3, 6-7 August 1986; dimensions in $\text{kg m}^{-2} \text{s}^{-1}$. Currents were measured with Endeco Model 174LR current meters. Water samples for suspended sediments were sampled with a pump and filtered at sea.

migrated between the 3-m and 6-m isobaths with flooding and ebbing tide respectively. The highest surface turbidities occurred landward of the front. At the deeper stations, surface concentrations were consistently less than 0.10 kg m^{-3} and near-bottom concentrations only exceeded 1 kg m^{-3} for a single 4-h period (5-6 August 1986).

The 1986 results showed seaward sediment fluxes over the delta front to be dominated by gravity-driven hyperpycnal plumes (underflows) which descended at plunge-point fronts near the edge of the shallow (<5 m) plateau-like delta front top. Although we had initially expected the underflows to be locally concentrated or channelized, this was not the case; acoustic water-column profiles at multiple frequencies (27, 123, 200 kHz) showed that the plumes associated with the summer flood season were widespread. Owing to mixing by tidal currents over the top of the delta front, salinities in the underflows were only slightly lower than those of the overlying ambient water. During the 1986 cruise, these underflows typically deposited most of their suspended sediment before passing beneath the vessel moored at the deeper (~10 m) stations. The most prominent underflow activity was observed at two stations located near the central plume axis which were occupied simultaneously for 26 h on 6-7 August 1986. The stations were separated by a horizontal distance of 1,300 m. Time series of instantaneous depth-varying sediment fluxes at the shallow station show that the reversing along-slope sediment flux maxima were roughly twice as large as the downslope components. But the cross-slope fluxes (Fig. 1) were exclusively seaward (downslope) and thus resulted in

a greater net transport. The highest seaward fluxes occurred in the lower 1-2 m of the water column between the times of maximum flood and maximum ebb (Fig. 1). Similar patterns were observed at the other two shallow stations.

Details of the vertical distribution of suspended sediment concentration, C_s , and excess bulk density σ ($\sigma = \rho - 1,000$ where ρ is total bulk density of the water/sediment mixture in kg m^{-3}), as observed during the peak and near the end of an underflow event (at 08:00 hours and 12:00 hours on 7 August 1986), are shown in Fig. 2. The index, σ_t , of the excess density of clear water is also indicated. In the cases shown, suspended sediment concentrations of 10 kg m^{-3} , and corresponding excess densities, produced a dilute hyperpycnal layer about 1 m thick. Greater thicknesses (up to 4 m) were observed at other times. We consider that the downslope advection within the hyperpycnal plume of mixed, but lower-salinity, water from the vicinity of the river mouth caused the vertical instability in σ_t in the bottom layer. This effect was a consistent feature of the hyperpycnal plumes observed during the 1986 and 1985 (ref. 5) cruises. Stable stratification was maintained by virtue of the high suspended-sediment concentrations; however, under such circumstances the deposition of suspended sediment must ultimately cause vertical instability near the top of the plume.

At the shallow station, net time-averaged (25 h) downslope sediment flux (Fig. 3) significantly exceeded net isobath-parallel flux, and was concentrated within 2 m of the bed (Fig. 3a). At the deeper station (Fig. 3b) the total suspended sediment flux was markedly less and no underflow was apparent. The depth-integrated, time-averaged vector resultants of net sediment flux at the shallow and deep stations were respectively $1.42 \text{ kg s}^{-1} \text{ m}^{-1}$ towards 146° and $0.54 \text{ kg s}^{-1} \text{ m}^{-1}$ towards 142°; both resultants were roughly downslope. Considering that the stations were only 1,300 m apart, this represents a steep gradient in sediment flux, implying an average deposition rate over the delta-front slope of 58.5 kg m^{-2} per day. Settling velocities of the suspended sediments were sufficient to permit this.

Significantly, stations occupied during calm weather in October 1987 revealed no underflows or hyperpycnal layers at either the shallow (~5 m) or deep (~10 m) stations. Strong, tide-driven isobath-parallel transport prevailed. The 1987 data, obtained at a time of negligible discharge, suggest that tidal stirring alone is probably insufficient to nourish widespread underflows such as we observed in 1986; such underflows seem to require a continual fluvial supply of suspended sediment. Storm-wave-induced resuspension of delta front muds appears, on the other hand, to be highly effective in generating dense, but episodic, hyperpycnal plumes. An intense northerly storm which persisted for five days during the 1987 cruise produced waves with significant heights of over 4 m. Water samples obtained in shallow water (<5 m) during the storm, and in deeper water (12 m) immediately after the storm, revealed near-bottom layers of fluid mud about 1.0-1.2 m thick. Suspended sediment concentrations observed at ~0.8 m above the bed at

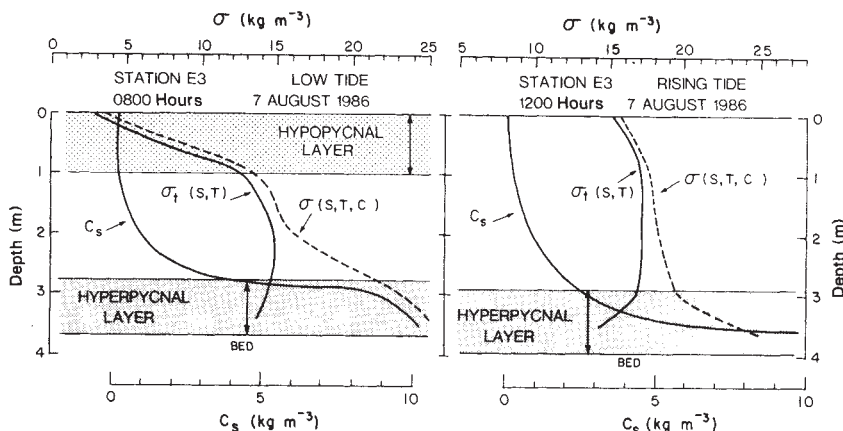


Fig. 2 Vertical profiles of suspended sediment concentration (C_s), excess density (σ_t) due to salinity and temperature only, and excess density (σ) due to the combined effects of salinity, temperature, and C_s at Station E3, 6-7 August 1986 during the middle (08:00) and end (12:00) of an underflow event. Salinity and temperature measured *in situ* with an InterOcean CTD system.

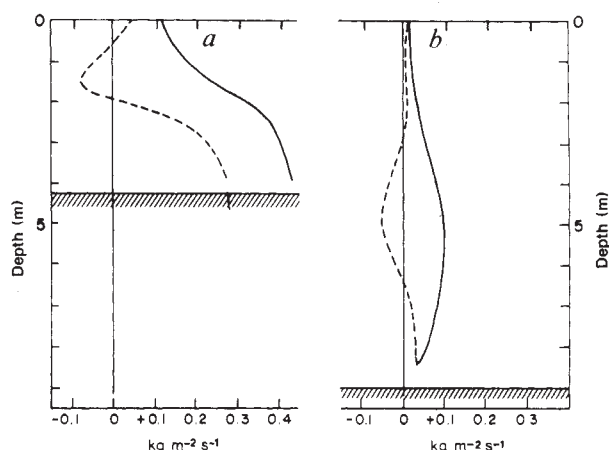


Fig. 3 Depth-dependent sediment fluxes, time-averaged over a diurnal tidal cycle (25 h; 6–7 August 1986) at *a*, shallow (E3) and *b*, deep (E11) stations. Solid line indicates cross-slope component; dashed line indicates isobath-parallel component. The two stations were occupied simultaneously and were separated by a horizontal distance of 1,300 m.

a depth of 12 m on 17 October 1987 averaged 252 kg m^{-3} . The fluid mud layers, identifiable in the post-storm phase as 'white' areas of no acoustic return on side-scan sonar imagery, appeared (from sonargraphs) to be confined to depressions on the slope and to be ponded at the base of the slope.

A simple model for describing the balance of forces in a gravity-driven density plume has the form

$$-(\tau_b + \tau_t) = \rho_l U^2 (K_b + K_t) = \rho_l g' h' \sin \beta \quad (1)$$

in which τ_b and τ_t are respectively the shear stresses at the bottom and top of the plume, ρ_l is the bulk density of the plume, which is greater than the density of the ambient water, ρ_a , U is underflow velocity downslope, K_b and K_t are respectively the drag coefficients applicable to the bottom and top of the plume, h' the vertical thickness of the underflow, β is bed slope which over the Huanghe delta front slope was 0.6° , and g' is reduced gravity: $g' = g(\rho_l - \rho_a)/\rho_a$ where g is acceleration due to gravity. For the 10 sets of hourly measurements during which underflows were obviously operating at station E3 in 1986, average values were: $\bar{U}_1 = 0.15 \text{ m s}^{-1}$; $\bar{h}' = 1.06 \text{ m}$; $\bar{g}' = 0.095 \text{ m s}^{-2}$; $|\tau_b + \tau_t| = 1.08 \text{ Pa}$ and thus $(K_b + K_t) = 0.046$. The instantaneous values of the combined drag coefficients ($K_b + K_t$) observed during the pronounced underflow event near slack tide on 7 August 1986 (07:00–08:00 hours) ranged from 0.011 to 0.03; these values are probably more appropriate than the overall mean. From the measured bulk densities and h' values associated with the storm-induced underflows in October 1987, we estimate g' to have been $\sim 1.49 \text{ m s}^{-2}$ and $|\tau_b + \tau_t|$ to have been $\sim 18.26 \text{ Pa}$. If it is assumed, on the basis of the 1986 results, that $(K_b + K_t) \approx$

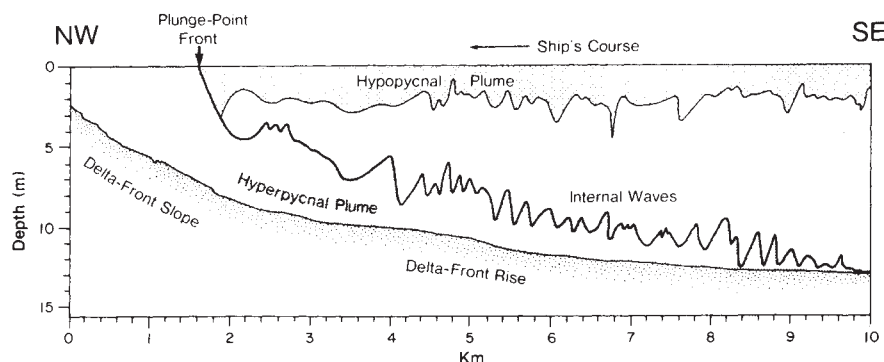
0.01–0.04, then the minimum corresponding storm-induced underflow velocities U_1 would have been somewhere between 0.62 m s^{-1} and 1.24 m s^{-1} . The associated downslope sediment fluxes would have been between $156 \text{ kg s}^{-1} \text{ m}^{-1}$ and $312 \text{ kg s}^{-1} \text{ m}^{-1}$. These values exceed the flux measured in 1986 by two orders of magnitude. It is likely that C_s and h' during the storm were greater than the values we observed after the storm; hence U_1 and sediment flux may have been greater.

Drag coefficients, K_b at the bases of turbidity currents have been estimated to lie in the range 0.0035–0.0050 (ref. 3). Our estimates from the 1986 data of total drag coefficient ($K_b + K_t$) values of 0.01–0.05 suggest that the observed rapid decay of dilute hyperpycnal plumes and consequent deposition of silt cannot be explained simply in terms of bottom drag. Furthermore, at least during the 1986 observations, the rate of deposition greatly exceeded the rate of resuspension: the dilute plumes were not autosuspending. Appreciable drag must have prevailed at the pycnocline separating overlying water from the underflow. The strong tidal currents could have contributed to vertical momentum exchange. Once deposition began, instability and enhanced mixing near the top of the plume would have resulted from the salinity deficit within the underflow. Another process observed at the top of the underflows was the activity of high frequency ($2.5\text{--}5.0 \times 10^{-3} \text{ Hz}$) internal waves of relatively large amplitude (Fig. 4). These internal waves were first identified during the 1985 cruise¹¹ and were found to have frequencies near the local Brunt–Väisälä frequency ($3.3 \pm 0.8 \times 10^{-3} \text{ Hz}$). During the 1986 cruise we found them to be ubiquitous features, often present at the upper interface of the hyperpycnal plumes; they probably contributed to plume deceleration.

We conclude that the dilute but laterally widespread and frequently recurring underflows, such as we observed in 1986, are the chief cause of accretion of the modern Yellow River's delta-front slope. Denser, storm-induced underflows probably act together with silt flows and other mass-movement mechanisms to deposit sediment on the delta front rise at the base of the slope. None of these processes appear to transport sediment very far afield. Notably, successive surveys of the subaqueous Yellow River delta¹² show that the shallow ($<10 \text{ m}$) isobaths are prograding at rates in excess of 1 km yr^{-1} . Also, about 90% of the sediment load debouched by the Yellow River since 1976 has remained within 30 km of the mouth¹⁰. We consider the deposition of silts from hyperpycnal plumes to be responsible for this rapid and spatially concentrated accumulation. In contrast to the dispersal and deposition of fine sediments off the mouths of rivers such as the Amazon⁶, where buoyant surface plumes transport suspended sediments considerable distances, deposition over the modern subaqueous Yellow River delta is immediate. The processes involved may have been more common in the geological past.

This work was supported by the US Office of Naval Research (Geology and Geophysics Program), the Geological Survey of Canada, and the Shandong College of Oceanography.

Fig. 4 Profile of hyperpycnal (dense) and hypopycnal (buoyant) plumes, the associated plunge-point front, and internal waves at the top of the hyperpycnal plume, as observed during an acoustic transect surveyed on 13 August 1986 over the delta front. The figure has been redrawn from a 200-kHz Ross echo sounder record. The traverse was somewhat oblique to the isobaths.



Received 24 July 1987; accepted 27 January 1988.

1. Pallesen, T. R. *Turbidity Currents*. Institute of Hydrodynamics and Hydraulic Engineering, Technical University of Denmark, Series Paper 32, 115 pp (1983).
2. Siegenthaler, C. & Buhler, J. *Mar. Geol.* **64**, 19–40 (1985).
3. Komar, P. D. in *The Sea: Marine Modelling*, Vol. 6 (eds Goldberg, E. D., McCave, I. N., O'Brien, J. J. & Steele, J. H. 603–621 (Wiley, New York, 1977).
4. Seymour, R. J. *Ocean Engng* **13**, 435–447 (1986).
5. Wright, L. D. *et al. Geo. Mar. Lett.* **6**, 151–174 (1986).
6. Curtin, T. B. & Legeckis, R. V. *Cont. Shelf Res.* **6**, 31–51 (1986).
7. Keller, G. H. & Prior, D. B. *Geo. Mar. Lett.* **6**, 63–120 (1986).
8. Qin, Y. & Li, F. *Proc. Int. Symp. on Sedimentation on the Continental Shelf with Special Reference to the East China Sea* **1**, 91–101 (Hangzhou China, 1983).
9. Ren, M. & Shi, Y.-L. *Cont. Shelf Res.* **6**, 785–810 (1986).
10. Bornhold, B. D. *et al. Geo. Mar. Lett.* **6**, 77–83 (1986).
11. Wright, L. D. *et al. Geo. Mar. Lett.* **6**, 115–120 (1986).
12. Yang, Z.-S. & Lu, Z. *Eos* **68**, 1735 (1987).

The structure of the Rockall Trough from seismic refraction and wide-angle reflection measurements

D. G. Roberts, A. Ginzberg, K. Nunn & R. McQuillin*

British Petroleum, Britannic House, Moor Lane,
London EC2Y 9BU, UK

* Britoil, 301 St Vincent Street, Glasgow G2 5DD, UK

There are few detailed descriptions of the deep crustal structure of failed cratonic rifts and passive margins, despite their importance in constraining the variety of structural models invoked to account for their common structural symmetry, or gross half-graben form^{1–3}. Such deep crustal-structure profiles also have considerable value in defining the continent/ocean boundary, constraining plate reconstructions, and determining the overall setting of sedimentary basin development. The Rockall Trough in the North Atlantic Ocean is a sinuous elongate bathymetric depression ~200 km wide and more than 2,000 m deep lying between Rockall Plateau and the British Isles, and the structure, nature and age of its underlying crust are disputed. In the absence of deep crustal-structure profiles, the presence of oceanic crust has been assumed because the trough has to be accommodated in reconstructions of the North Atlantic, although stretched continental crust is equally plausible. The nature of the crust is important for the precise reconstruction of the late Jurassic/early Cretaceous plate history and setting of sedimentary basins of the North Atlantic^{4–8}. The nature of the crust beneath the trough as well as the structure, thickness and composition of the sedimentary cover is also relevant to oil exploration. Here we report the first observations of deep crustal structure across the Rockall Trough and its margins from seismic refraction and wide-angle reflection profiles. Our results show that the northern Rockall Trough is underlain by highly stretched continental crust, and not by oceanic crust as previously postulated.

The structure of the crust in the areas bordering the Rockall Trough to the east and west is well known. Under England and Scotland there is a three-layer continental crust⁹ comprising a 6.1–6.2 km s⁻¹ upper crust, a 6.5 km s⁻¹ mid-crust with a 7.0 km s⁻¹ lower crust overlying an 8.0–8.1 km s⁻¹ layer. North and west of Scotland, a two-layer continental crust with a 6.1 km s⁻¹ upper crust, 6.5 km s⁻¹ lower crust and 8.0 km s⁻¹ Moho phase is present and thins westward^{10–13}.

Scrutton¹⁴ has reported a two-layer continental crust beneath Rockall Plateau comprising a 6.35 km s⁻¹ upper crust and a 7.02 km s⁻¹ lower crust overlying an 8.2 km s⁻¹ upper mantle. But the crustal structure has remained ambiguous within the Rockall Trough. From regional considerations and refraction data¹⁶ it was¹⁵ inferred that the Rockall Trough was probably underlain by oceanic crust but refraction profiles obtained in the trough are not uniquely interpretable in terms of crustal structure¹⁷. Thus oceanic and stretched continental crust models

have been variously favoured on the basis of poor data sets^{4,13,15,18}. Part of this difficulty is caused by the presence of pervasive intrusives and extrusives of early Tertiary age. More recently, an oceanic structure has been proposed from preliminary interpretation of a wide-angle reflection profile along the axis of the trough¹⁹.

Our observations of deep crustal structure across the Rockall Trough and its margins involve seismic refraction and wide-angle reflection profiles (Fig. 1). Data quality on the two profiles varied widely from fair to poor. The best data were recorded at the shelf and slope ends of both profiles. In these locations some good basement (P_g) arrivals were recorded, as well as strong P_1P (mid-crustal discontinuity) and P_mP (Moho) reflected phases (Fig. 2). The quality of the recorded arrivals deteriorated very rapidly once deep water was reached. Thus, the ocean-bottom seismometers (OBS) in the central portions of the profiles recorded refracted and reflected arrivals from the sedimentary cover, followed by P_g arrivals, which died out at very short ranges. In some cases P_g later arrivals and later reflected phases can be recognized (Fig. 2). As a result of the high noise level at ranges above 30 km, no P_n arrivals were observed on either line. The poor transmission properties in the Rockall Trough have been noted by earlier investigators¹⁴. They can be ascribed to the inhomogeneity of the pre-Tertiary sedimentary cover which is pervasively intruded and overlain by a considerable thickness of basalt flows¹⁸.

The interpretation consisted of the computation by standard techniques of the true refractor velocities from the refracted arrivals, and of the average velocities to the deep reflectors. The simple velocity-depth models thus obtained were used as the initial input models for forward modelling by ray tracing. At the extremities of the profiles, the first models were based on published results^{9,11,13,14}. The ray-tracing package used was SEIS81 (ref. 20).

Synthetic seismograms were computed for selected OBS positions to check whether the general distribution of energy for the interpreted model matched the observed data. We computed both ray theoretical seismograms²⁰ and reflectivity seismograms^{21,22}. To obtain a single model that satisfied all OBS positions, the input model was iteratively adjusted at each OBS position until a satisfactory travel-time fit was obtained. This process was then followed by iterations between OBS positions. Exemplary models, ray diagrams and record sections are shown in Fig. 2 and 3.

The crustal structure shown in Fig. 4 has been derived from analysis and modelling of the data from the OBS positions shown. Analysis of the computed refraction velocities shows three similar velocity layers between the sea bed and the crystalline basement on profiles 2 and 5 (Fig. 4). These layers are thought to represent Miocene/Recent, late Cretaceous/Eocene, basalt and/or early Mesozoic sediments respectively¹⁵. Profile 5 also shows evidence in its easternmost part of an additional 5.3 km s⁻¹ layer, thought to represent late Palaeozoic, early Mesozoic or fractured Torridonian sediments¹¹. On both profiles, the crystalline basement exhibits a velocity that varies from 6.0 to 6.25 km s⁻¹. The average crustal velocity calculated from P_mP reflections lies in the range 6.2–6.5 km s⁻¹. For modelling purposes, velocities of 6.6 km s⁻¹ and 8.0 km s⁻¹ were adopted for the top of the lower crust and upper mantle respectively.

In more detail, the interpreted model of profile 2 shows that the western flank of the Rockall Trough consists of a dipping basement having a velocity of 6.0 km s⁻¹, covered by two layers of sediments which thicken gradually eastward. The P_mP and P_1P reflections recorded here indicate a gradual eastward thinning of the crust from a thickness of ~28 km to ~23 km at 20 km distance along the profile (d). At 40 km (d), the thickness of the sedimentary sequence increases rapidly and an additional layer with a velocity of 4.7 km s⁻¹ appears. From 50 km (d) to the eastern flank of the trough, the velocity layering of the

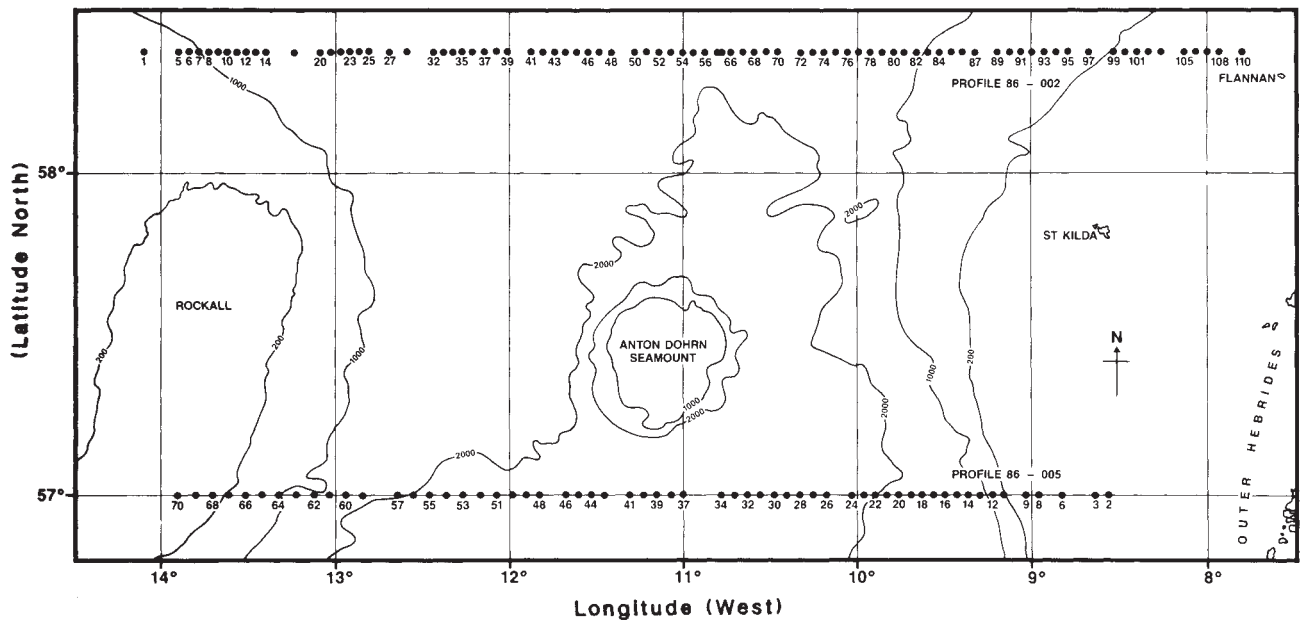


Fig. 1 Bathymetric map of the Rockall area showing the position of the refraction profiles.

Methods. The two refraction and wide-angle reflection profiles were acquired and processed for BP Exploration and Britoil by the Institute of Geophysics, Hamburg University using RV *Valdivia* in February–March 1986. The profiles were shot using closely spaced dynamite shots in three or four overlapping segments with 40 OBS deployed along each. Small (25–50 kg) explosive charges could only be used due to severe weather conditions.

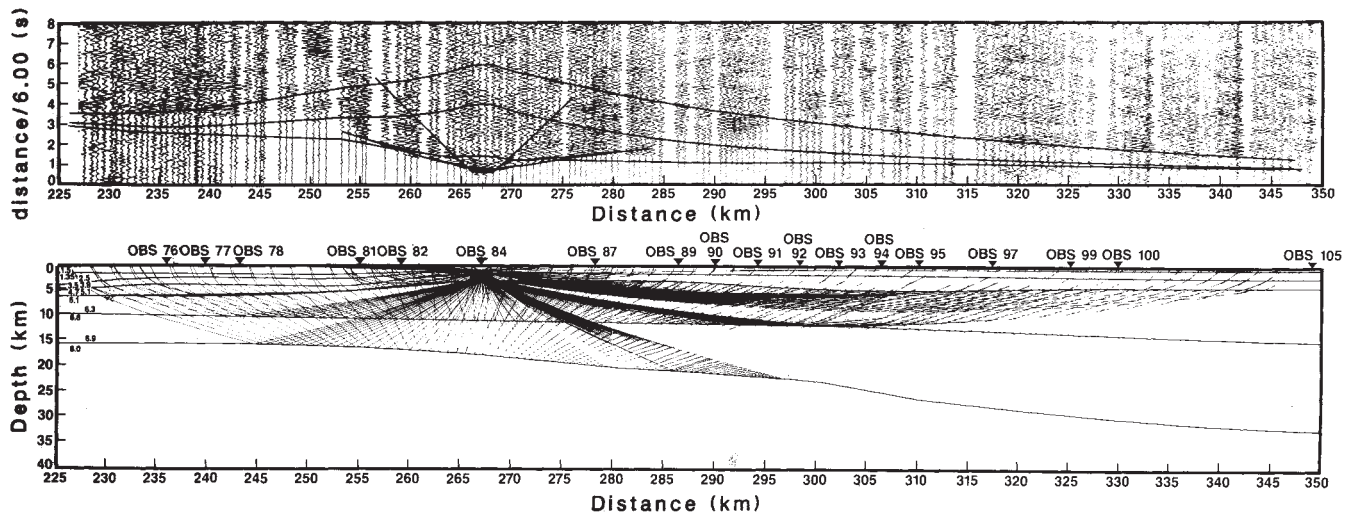


Fig. 2 Model, ray diagram and record section of OBS 84, profile 5. Note the strong reflections on either side of the OBS.

Methods. During processing of refraction and wide-angle reflection profiles, the four channels (three seismic and a timing channel) were amplified, filtered by a low-pass anti-aliasing filter and digitized at an 8-ms sampling rate. The digitized data were filtered (3–12 Hz band pass), deconvolved, normalized and plotted as record sections with a reduction velocity of 6.0 km s^{-1} . Selected record sections were plotted with real amplitudes for amplitude modelling with synthetic seismograms.

sedimentary cover is maintained with slight local variation. The mid-crustal discontinuity reflection (P_1P) can be clearly seen at OBS 17 (Fig. 3), and was used to determine the velocity to the discontinuity and its depth. It should be noted here that a slightly higher velocity of 6.2 km s^{-1} was recorded in this part of the trough. Undulations and thickness variations in the shallow velocity layers between 50 and 150 km (d) may indicate underlying basement structure. The continued presence of the mid-crustal discontinuity is inferred from the P_1P reflection seen at several OBS positions across the central portion of the profile. The 2.0 km s^{-1} layer thickens gradually westward, and its base is at a depth of over 4 km at 200 km (d), whereas the 3.5 km s^{-1} interval remains at about the same level with some broad undulations. Eastwards from 250 km (d), the data quality improves and strong reflections are recorded as well as much better basement arrivals. The crystalline basement shoals and the over-

lying sedimentary layers thin in two broad undulations, probably reflecting the underlying basement tilted blocks. The 2.0 km s^{-1} layer thins rapidly eastwards and wedges out completely at 275 km (d). The 3.5 km s^{-1} layer maintains a constant thickness while the depth to basement decreases. East of 275 km (d), a basin containing over 4 km of sediments, possibly Mesozoic, is present. Comparison of the east and west margins of the Trough shows that the Moho is shallowest adjacent to the east margin and dips gently westward, emphasizing an overall asymmetry to the structure of the trough.

Profile 5 shows a similar velocity sequence and structure. The east flank of the Rockall Plateau is characterized here by a 4.5 km s^{-1} layer overlying the 6.0 km s^{-1} basement. The 3.4 km s^{-1} and the 2.3 km s^{-1} layers appear further east. The sedimentary cover thickens gradually to 40 km (d), beyond which there is an abrupt increase in thickness. The mid-crustal

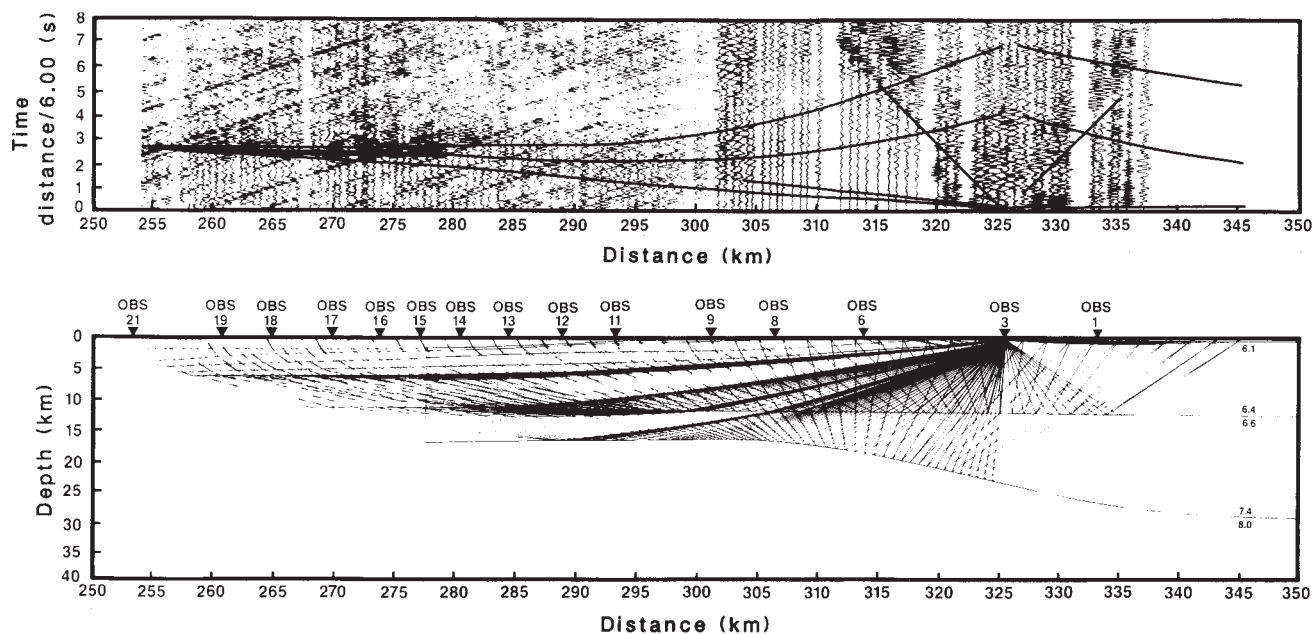


Fig. 3 Model, ray diagram and record section of OBS 6, profile 5. Note the closely spaced P_mP and P_iP reflections.

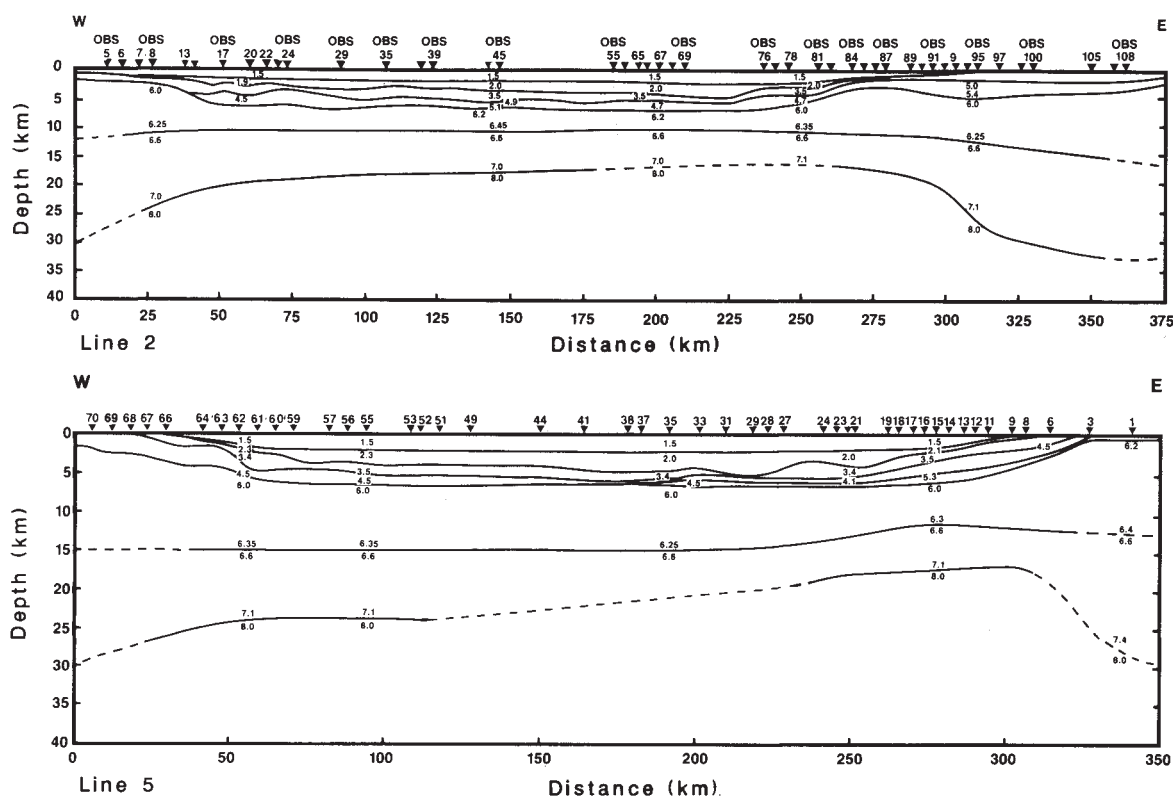


Fig. 4 Crustal cross-sections of profiles 2 and 5. OBS positions used in the construction of this model are shown at the top. Also shown are the main velocity values. Interfaces shown in dashed lines are unconstrained.

discontinuity is at a depth of some 15 km in the west and shoals gradually to the east, whereas the Moho is at a calculated depth of 28 km in the west and 25 km at 50 km (d) to the east. Good P_mP and P_iP reflections were recorded to 110 km. They indicate a continuing shoaling of the Moho while the depth of the mid-crustal discontinuity remains at 15 km. Further east, the 2.3 km s^{-1} layer thickens considerably. The velocity of this layer is lower in the centre of the trough (2.0 km s^{-1}). Good P_iP reflections were recorded at 180 km (Fig. 3). Eastwards from

180 km (d), the 2.0 km s^{-1} layer begins to thin in three broad steps. The underlying sedimentary layers also shoal, and an additional layer having a velocity of 5.3 km s^{-1} was recognized here immediately above the crystalline basement. On the eastern flank of the trough, strong deep reflections were again recorded, so that the depth to the deep refractors is well-constrained along this part of the profile. The sedimentary cover, which is still quite thick at 250 km (d), thins gradually eastward. The maximum thinning of the crust is interpreted as being at 300 km

(d), and an eastward thickening of the crust is evident from this position (Fig. 4). The sedimentary cover on the slope to the east is very thin, or almost completely missing. Profile 5 again shows a comparable, but stronger, asymmetry in crustal structure to profile 2.

Our results demonstrate that the crust underlying the 6 km of water and sediment over in the Rockall Trough is considerably thinner under the trough than under the known continental areas lying both to the east and west^{9,11,13,14}. It has been shown that the thinned crust is a two-layer crust along the entire length of both profiles, with a mid-crustal discontinuity separating an upper crust from a lower crust with a considerable velocity gradient (Fig. 4). These results make a definite determination of the nature of the crust underneath the Rockall Trough possible. This is because the results of seismic experiments over the oceanic crust in many worldwide locations have generated an accepted definition of the oceanic crust^{23,24}: an oceanic crust is typified by a thin 5.5 km s⁻¹ basaltic layer (layer 2) overlying a high velocity (6.5 km s⁻¹), low-gradient crust 5–10 km thick (layer 3). By contrast the continental crust, normal or stretched, is composed of a 6.0 km s⁻¹ upper crust and a 6.5 km s⁻¹ lower crust with a moderate-to-strong velocity gradient²⁵.

As a two-layer crust with a basement velocity of 6.0–6.2 km s⁻¹ and a considerable velocity gradient in the lower crust was detected underneath both of our profiles, we have shown unequivocally that the Rockall Trough is underlain by a continental crust which was asymmetrically thinned by stretching. The amount of stretching observed in the Rockall Trough is similar to that observed in the Bay of Biscay²⁶, where a tilted fault block structure for the upper crust is suggested by the relief in the lower sedimentary layers²⁶. Further, the pattern of stretching in the upper and lower crust observed in the western part of the Rockall Trough and Biscay is similar. In Biscay, the data²⁶ show that the mid-crustal discontinuity or brittle ductile boundary (interpreted from reflection and refraction data) rises towards the continent-ocean boundary and as the overlying basement involved listric normal faults flatten. In the Rockall Trough, the mid-crustal discontinuity rises to reach its minimum depth closer to the thicker crust beneath the east margin of the Rockall Trough and above the shallowest elevation of the Moho, defining the maximum stretching of the crust. The similarity of stretching patterns and structural asymmetry of the Rockall Trough and the Biscay margin suggests that the Biscay margin may be equivalent to the completely rifted western part of the Rockall Trough. Besides documenting the continental structure of the crust beneath Rockall Trough, the results also provide a unique demonstration of the asymmetry in deep crustal structure invoked for many failed rifts and passive margins derived from studies of shallow seismic data.

We thank the Board of British Petroleum and Britoil for permission to publish the work, Professor J. Makris, the crew of RV *Valdivia* and staff at Hamburg University for their help. Discussions with P. A. Vingoe, I. R. Vann, S. A. D. Bamford, N. Gillette, M. Earle and R. Rosenthal were valuable. B. J. Gooden and C. C. Swaine provided technical assistance.

Received 1 September 1987; accepted 2 March 1988.

- Bally, A. W. *Phil. Trans. R. Soc. A* **305**, 325–328 (1982).
- Wernicke, B. *Can. J. Earth Sci.* **22**, 108–125 (1985).
- Lister, G. *et al. Geology* **14**, 246–250 (1986).
- Talwani, M. and Eldholm, O. *Geol. Soc. Am. Bull.* **88**, 969–999 (1977).
- Kristoffersen, Y. *Earth Planet. Sci. Lett.* **38**, 273–290 (1978).
- Srivastava, S. P. *Geophys. J. R. astr. Soc.* **53**, 259–287 (1978).
- Srivastava, S. P. *Tectonophysics* **114**, 29–53 (1985).
- Srivastava, S. P. & Tapscott, C. R. in *The Geology of North America* Vol. M, 379–404 (Geol. Soc. Am., 1986).
- Bamford, D., Nunn, K., Prodehl, C. & Jacob, B. *Geophys. J. R. astr. Soc.* **54**, 43–60 (1978).
- Bott, M. H. P. & Smith, P. J. *Geophys. J. R. astr. Soc.* **76**, 383–398 (1984).
- Jones, E. J. W., White, R. S., Hughes, V. J., Matthews, D. H. & Clayton, B. R. *Geophysics* **49**, 1605–1621 (1984).
- Armour, A. R. thesis, Univ. Durham (1977).
- Bott, M. H. P., Armour, A. R., Himsworth, E. M., Murphy, T. & Wylie, G. *Tectonophysics* **59**, 217–231 (1979).
- Scrutton, R. A. *Geophys. J. R. astr. Soc.* **27**, 259–275 (1972).

- Roberts, D. G. *Phil. Trans. R. Soc. A* **278**, 447–500 (1975).
- Ewing, M. & Ewing, J. *Bull. geol. Soc. Am.* **50**, 291–317 (1959).
- Jones, E. J. W., Ewing, M., Ewing, J. I. & Eittrheim, S. L. *J. geophys. Res.* **75**, 1655–1681 (1970).
- Roberts, D. G., Bott, M. H. P. & Uruski, C. in *Structure and Development of the Greenland–Scotland Ridge* (eds Bott, M. H. P., Saxon, E., Talwani, M. & Thiede, J. K.) 133–158 (Plenum, New York and London, 1983).
- White, R. S. *et al. Abstr. Geol. Soc. Newsletter* **16**(2), 93 (1986).
- Cerveny, V. & Psencik, J. *2-D Seismic Ray package* (Charles University, Prague, 1981).
- Fuchs, K. & Muller, G. *Geophys. J. R. astr. Soc.* **23**, 417–433 (1971).
- Sandmeier, K. J. & Wenzel, F. *Geophys. Res. Lett.* **13**, 22–25 (1986).
- Spudich, P. & Orcutt, R. *Geophys. Space Phys.* **18**, 627–645 (1980).
- White, R. S. in *Ophiolites and Oceanic Lithosphere* (eds Gacs, I. G., Lippard, S. J. & Shelton, A. W.) 101–111 (Blackwell, Oxford, 1984).
- Prodehl, C. *Structure of the Earth's Crust and Upper Mantle* Vol II/11a, 97–206 (Springer, Heidelberg, 1984).
- Avedik, F. *et al. Phil. Trans. R. Soc. A* **305**, 5–25 (1982).

The dielectric constant approach to speciation and ion pairing at high temperature and pressure

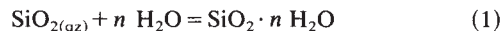
John V. Walther* & Jacques Schott†‡

* Department of Geological Sciences, Northwestern University, Evanston, Illinois 60201, USA

† Laboratoire de Minéralogie et Cristallographie, Université Paul Sabatier, 31062 Toulouse, France

Fluids in the Earth's crust commonly contain, in addition to the solvent, H₂O, high concentrations of other volatiles (such as CO₂, CH₄ or N₂) and/or salts (such as NaCl, KCl or CaCl₂). The solubility of minerals in these fluids is of prime importance for understanding metasomatic processes. Many theoretical and empirical relations have been proposed to account for various aspects of solubility behaviour and speciation. Here we show that quartz solubility behaviour in solutions where the activity of H₂O is reduced from unity is easily modelled from knowledge of changes in the dielectric constant of the solvent. The approach may be generalized to all neutral aqueous species. We further demonstrate the usefulness of considering a modified form of the Fuoss equation¹ for the calculation of the dielectric-constant dependence of ion pairing at high temperatures and pressures. This approach thus allows the prediction of speciation in complex fluids at high temperature and pressure for reduced activities of H₂O.

Quartz solubility in pure H₂O is well known to temperatures of 900 °C and pressures of 10 kbar (ref. 2). At temperatures and pressures above the critical point of H₂O, the addition of CO₂ or Ar dramatically decreases quartz solubility^{3–6}. This decrease has been ascribed to the reduction in activity of H₂O ($a_{\text{H}_2\text{O}}$) according to the reaction:



where $\text{SiO}_2 \cdot n \text{H}_2\text{O}$ stands for the neutral hydrated species in solution. Analysis of experimental quartz solubilities according to equation (1) leads to values of n , the hydration number of aqueous silica, of ≥ 4 , rather than 2 as implied by writing H_4SiO_4 , which is generally regarded as the major silica species in mildly alkaline to acidic solutions^{3,4,6,7}. The immiscibility of mixtures of H₂O with CO₂ or Ar at subcritical conditions precludes their use to vary $a_{\text{H}_2\text{O}}$ at constant pressure and temperature. Considerable solubility data is, however, available on the quartz polymorph, amorphous silica, at 25 °C in aqueous salt solutions⁸ where $a_{\text{H}_2\text{O}}$ is decreased significantly below unity. In these solutions the solubility of amorphous silica decreases with increasing concentration of salt ('salting out'). An analysis similar to equation (1) shows n to vary with both concentration and the identity of the salt from almost zero to >4 (ref. 8). As temperature increases n decreases, so that the salt's effect on solubility decreases⁹. Judging from the limited data above the

‡ To whom correspondence should be addressed.

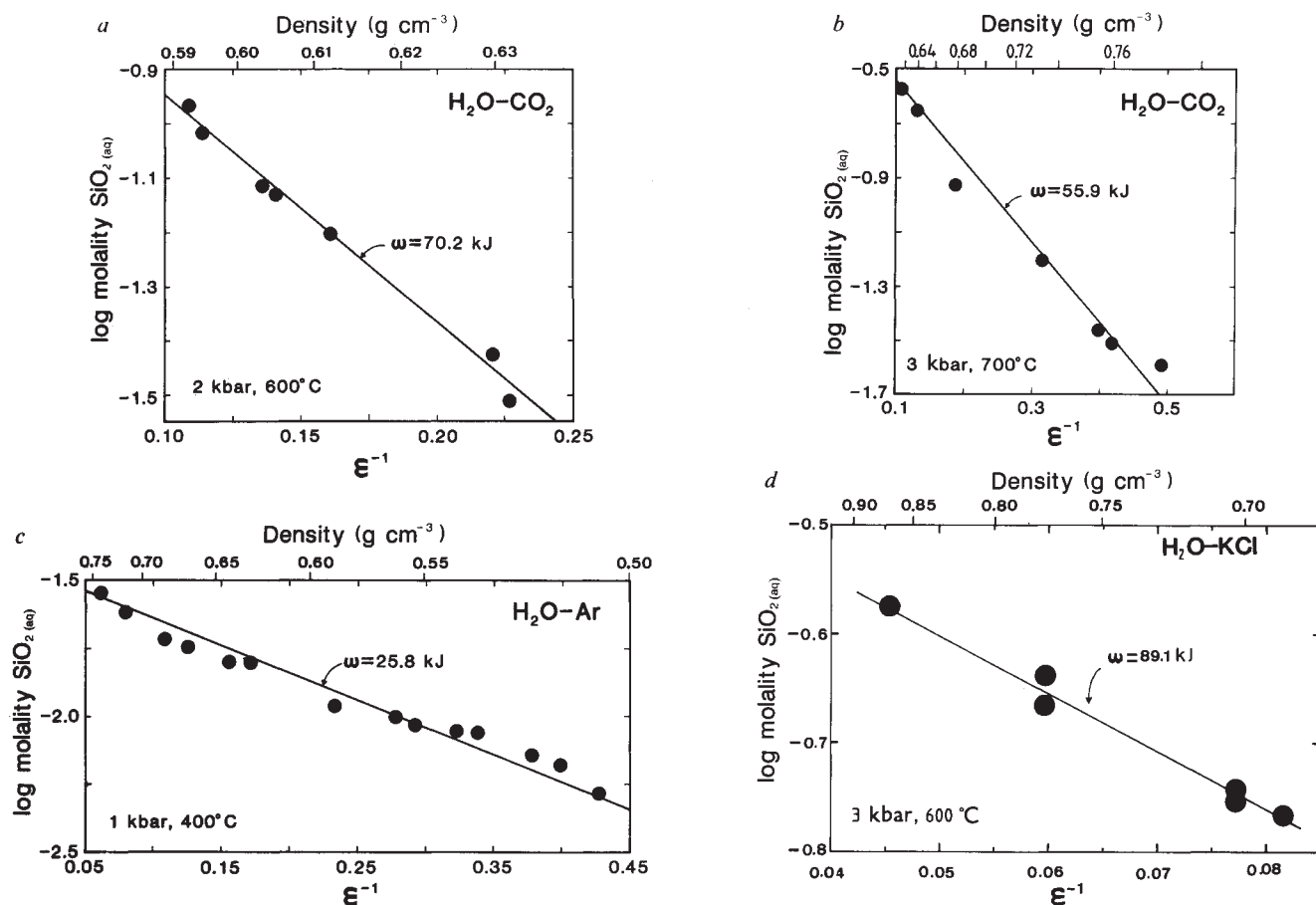


Fig. 1 Solubility of quartz expressed as log molality aqueous silica at the indicated temperature and pressure, as a function of the reciprocal of the dielectric constant of the fluid mixture: *a*, *b*, H₂O-CO₂; *c*, H₂O-Ar; *d*, H₂O-KCl. Solubility data are from refs 5 (Fig. 1*b*), 6 (Fig. 1*a*, *c*) and 10 (Fig. 1*d*). ω is the value of the Born interaction parameter per mole for the line indicated. For H₂O-KCl mixtures both the solubility and dielectric constant increase with increasing KCl. Density scales show that there is not a simple dependence of quartz solubility on fluid mixture density.

critical temperature of H₂O, the computed values of n decrease below unity ('salting in') for KCl and NaCl because the addition of salt increases the molality of SiO₂ in solution¹⁰. This behaviour has been interpreted as indicating the formation of significant ion pairing of alkalis with Si¹¹. Thus, if the behaviour of aqueous silica is any indication, the characterization of complex solutions at high temperatures and pressures requires dissociation constants for a variety of complexes unknown at lower temperature. In the case of solutions with reduced $a_{\text{H}_2\text{O}}$, the hydration number of each species must also be known. Such characterization is an almost hopeless task given the large number of species that may exist in crustal solutions and the variety of pressures and temperatures that need to be considered.

We propose an alternative approach which can explain all of the observations of aqueous silica behaviour without evoking a model of species hydration or appealing to alkali-Si complexes. Essentially the approach is an extension of the recent analysis of Dandurand and Schott^{12,13} of amorphous silica solubility in H₂O-organic and H₂O-salt mixtures at temperatures below the critical temperature of H₂O. They followed an electrostatic approach which takes into account SiO_{2(aq)}-solvent (Born¹⁴) and SiO_{2(aq)}-ion (Brønsted¹⁵) interactions to model the excess free energy properties of SiO_{2(aq)}. This can be expressed as follows:

$$\ln [m_w(\text{SiO}_2)/m(\text{SiO}_2)] = -(\omega/RT) (1/\epsilon_w - 1/\epsilon) + (A/\epsilon RT) (\sum c + \sum a) + (B/RT) (\sum c + \sum a) \quad (2)$$

where SiO₂ refers to the solvated silica species in solution, the

subscript w denotes pure H₂O values, R is the gas constant, T is absolute temperature, m is the molality of the species in parenthesis and ω is the Born parameter for SiO₂-solvent interaction. A and B relate the linear dependence of ϵ , the dielectric constant, to ionic strength, so that the last two terms of equation (2) account for the short-range interaction between SiO_{2(aq)} and cations, c , and anions, a , according to the Brønsted principle of specific interactions. In aqueous solutions of organic compounds Dandurand and Schott^{12,13} showed that the variation of $\ln[m(\text{SiO}_2)]$ with $1/\epsilon$ can be represented by a single linear equation independent of the nature of the organic compound, indicating that short-range interactions among neutral species are negligible. Hence, for organic-bearing systems, equation (2) reduces to

$$\ln [m_w(\text{SiO}_2)/m(\text{SiO}_2)] = -\omega/RT (1/\epsilon_w - 1/\epsilon) \quad (3)$$

Equation (2) should be equally valid at high temperatures and pressures; to apply it, dielectric properties must be known for the solution. The dielectric constant ϵ_0 of a pure fluid can be deduced from its molecular polarizability, its dipole moment and its molar volume by using the well-known Kirkwood equation¹⁶. Resulting values of ϵ_0 for Ar, CO₂, CH₄ and C₆H₆ at 400°C and 2 kbar are 1.28, 1.47, 1.40 and 1.94, respectively. The only measured value for a non-polar fluid under these conditions is for C₆H₆ and is reported to be 2.10 (ref. 17). The good agreement between the calculated and measured values of ϵ_0 for C₆H₆ indicates that the Kirkwood equation reliably calculates ϵ_0 for non-polar fluids at high temperatures and pressures. To compute ϵ for fluid mixtures, we elected to use the mixing equation of Looyenga¹⁸ because it predicts the dielectric properties of H₂O-C₆H₆ mixtures measured by Deul¹⁷ at

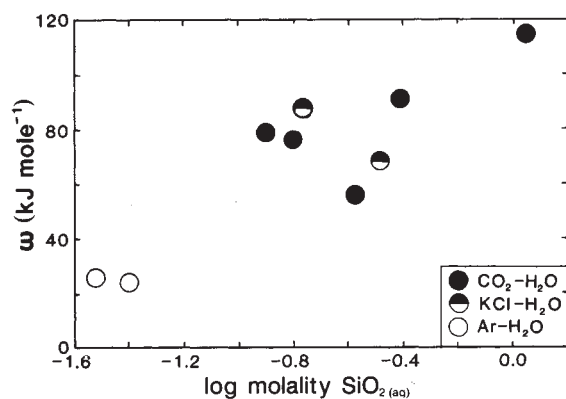


Fig. 2 Values of ω , the Born interaction parameter, for various mixtures as a function of log molality of aqueous silica in pure water.

400 °C to pressures as high as 3 kbar and therefore should give reasonably accurate values for H₂O mixtures with any non-polar fluid. Looyenga's formulation is a truncated Taylor expansion based on the permittivity of the fluid:

$$\epsilon = [\epsilon_o^{1/3} + \Phi_w(\epsilon_w^{1/3} - \epsilon_o^{1/3})]^3 \quad (4)$$

where the subscripts o and w stand for the non-polar fluid and H₂O, respectively, and Φ is the volume fraction in the mixture. We used the Kirkwood equation and equation (4) to compute ϵ for H₂O-Ar and H₂O-CO₂ mixtures. In the calculation we obtained volumes of the pure fluids and mixtures from a modified Redlich-Kwong equation¹⁹ and used the equation of Pitzer²⁰ to obtain g , the Kirkwood "orientation factor", for H₂O.

If, as discussed previously, the behaviour of quartz solubility is considered to be a function only of the dielectric constant at constant temperature and pressure, with no short-range interaction, then according to equation (3), plots of $\ln[m(\text{SiO}_2)]$ as a function of $1/\epsilon$ should be straight lines with slopes of $-\omega/RT$. Such plots for a number of pressures and temperatures are shown in Fig. 1.

We used a similar approach to analyse quartz solubility in H₂O-KCl mixtures. At high concentration and temperature, salts like NaCl and KCl exist in aqueous solutions primarily as neutral species. Unlike Ar or CO₂, these ion pairs are quite polar. We estimated the molar volume of KCl from the density of KCl fused salt, assuming that it was supercooled to 600 or 700 °C. We also assumed $g = 1$ and ideal mixing with H₂O. When ϵ is computed for H₂O-KCl mixtures, we find that the addition of KCl increases ϵ at high temperatures and pressures. Figure 1d shows quartz solubilities measured at 3 kbar and 600 °C in pure H₂O and high concentrations of KCl¹⁰ as a function of the calculated $1/\epsilon$; a linear dependence does exist. As with the increase in amorphous silica solubilities in aqueous solutions of formamide at 25 °C¹², the increase in SiO_{2(aq)} can be explained with an increase of ϵ in the solutions. Values of ω determined by this approach are plotted in Fig. 2 for mixtures with Ar, CO₂ and KCl as a function of the solubility of quartz in pure H₂O at the same temperature and pressure. Although some scatter exists, there does seem to be a small positive dependence of ω on $m_w(\text{SiO}_2)$. There may be some short-range interaction, the effect of which would obviously increase with increasing SiO_{2(aq)} concentration. Alternatively, ω may be increasing with increasing pressure and/or temperature, as represented by increasing SiO_{2(aq)} concentration. We conclude that the thermodynamics of neutral aqueous silica in H₂O mixtures with other fluids (such as CO₂ or CH₄) or high salt content (KCl, NaCl, CaCl₂, MgCl₂) may at least to a good approximation be characterized by equation (3) when changes in ϵ and a simple monomeric model for aqueous silica are considered. Similar behaviour should occur for other neutral species such as Al(OH)₃, H₂S, or any of the neutral ion pairs.

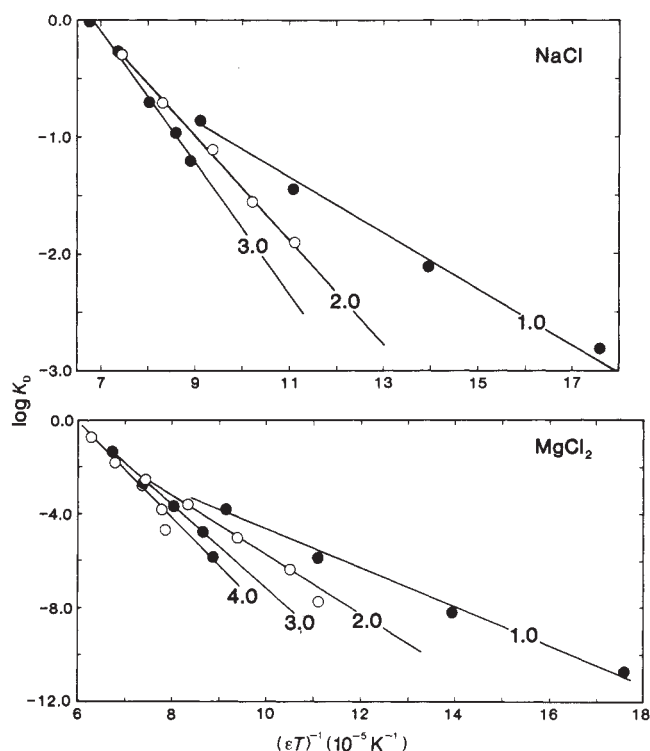


Fig. 3 The log of the dissociation constant of NaCl and MgCl₂ as a function of the reciprocal of the product of the dielectric constant and absolute temperature, at various pressures (curves labelled with pressure in kilobars).

It is tempting to extend the dielectric constant approach to the treatment of ion pairs. Since Bjerrum's early work²¹, many investigators have used an electrostatic approach to describe ion pair formation. As yet, however, this approach has not been successfully formulated for high-temperature, high-pressure solutions. For instance, the relationship derived by Fuoss¹ between ϵ and the dissociation constant, K_D for ion pairs:

$$\log K_D = 2.598 - 3 \log \bar{a} - 72,400/(\bar{a}\epsilon T) \quad (5)$$

(where $\bar{a}$ is the centre-to-centre distance of the ions in the pair, in ångströms) fails adequately to describe the pressure and temperature behaviour of K_D for any of the large number of salts in supercritical H₂O that have been measured. As a result, the electrostatic approach has been abandoned in favour of empirical methods based on changes in solvent density²²⁻²⁵.

The main problem with the application of the Fuoss equation, and with the electrostatic approach in general, is that it requires knowledge of the ion size parameter, $\bar{a}$. As has been discussed extensively, in the literature, $\bar{a}$ must account for all the short-range interactions that occurs, and as such can not be calculated *a priori* with any precision. In most treatments, such as the extended Debye-Hückel equation for the calculation of activity coefficients of charged species, $\bar{a}$ appears to be related to the surface charge potential of the ion rather than its actual size in solution. It is therefore reasonable to assume that $\bar{a}$ in equation (5) could take values that cannot be directly related to its physical size. Additionally, as ions are hydrated, it is likely that the $\bar{a}$ parameter for solvent interactions is not the same as for short-range interaction. We therefore introduce two different $\bar{a}$'s into the Fuoss equation, which thus becomes:

$$\log K_D = 2.598 - 3 \log \bar{a}_1 - 72,400/(\bar{a}_2\epsilon T) \quad (6)$$

Figure 3 shows $\log K_D$ for NaCl²⁶ and MgCl₂²⁷ as a function of $(\epsilon T)^{-1}$. As mentioned above the measured values do not fall on a single straight line as predicted by equation (5), but instead

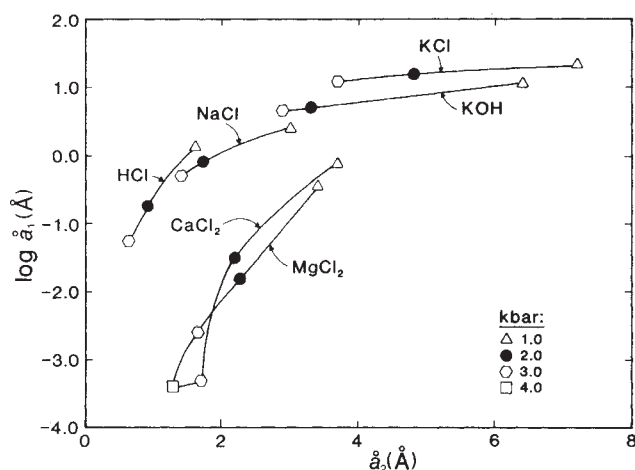


Fig. 4 Values of $\bar{a}_1$ and $\bar{a}_2$ obtained from fitting equation (6) to the dissociation constants of some representative salts at various pressures (kbar).

diverge as $\log K_D$ decreases. Note, however, that at constant pressure the value of $\log K_D$ does decrease linearly with $(\epsilon T)^{-1}$. All the salts we have checked, as well as H_2O dissociation, display similar behaviour. We have fitted $\log K_D$ at 1, 2 and 3 kbar with equation (6) to obtain values of $\bar{a}_1$ and $\bar{a}_2$. These values are plotted against each other in Fig. 4, together with those obtained from similar fits for HCl^{28} , KCl^{29} , KOH^{30} and $CaCl_2^{27}$. There is a smooth dependence of $\bar{a}_1$ on $\bar{a}_2$ for each electrolyte as a function of pressure. One can also show that the pressure dependence is proportional to the cation surface charge potential of each salt. We conclude that an electrostatic approach as outlined here can adequately describe the behaviour of ion pair formation as a function of temperature and pressure in the supercritical region.

Because the dielectric constant for mixtures of H_2O with species such as CO_2 , CH_4 , N_2 , $NaCl$, KCl and $CaCl_2$ can be approximated, use of equations (3) and (6) allows excess free energy and speciation calculations in a wide variety of crustal fluids. We may, therefore, be closer to characterizing metasomatic processes in complex fluid mixtures than was previously thought.

First Cretaceous mammal from India

G. V. R. Prasad* & Ashok Sahni†

* P. G. Department of Geology, University of Jammu, Jammu Tawi 180001, India

† Centre of Advanced Study in Geology, Panjab University, Chandigarh 160014, India

The record of Cretaceous mammals in particular and Mesozoic mammals in general is biased in favour of the Laurasian landmass¹. There are relatively few reports of these mammals from the southern continents representing Gondwanaland, and most of these are confined to northwestern South America^{2,3}. Here we report the discovery of a new eutherian mammal from the uppermost Maastrichtian sediments intercalated in a volcanic sequence in Naskal, Andhra Pradesh, India. This first record of a south-Asian Cretaceous mammal indicates a wide distribution of these creatures and suggests that the Indian plate was not isolated from northern continents despite the geological evidence. Our finds may also contribute to the understanding of the origin and evolution of therian mammals from mammal-like reptiles (therapsids) which are well known from the southern continents⁴.

The discovery of mammals from the Naskal intertrappean sediments is the result of a long search for Cretaceous-Palaeocene mammals in the Indian subcontinent⁵⁻⁷. Mammalian teeth were recovered along with remains of other vertebrates (fishes, frogs, lizards, snakes, turtles and crocodiles) from pinkish-white sandy marl of the intertrappean succession exposed 2 km north-east of the village of Naskal (Fig. 1). These sediments, which lie along the eastern fringe of the Deccan volcanics, have been dated on the basis of palaeontological, palaeomagnetic and radiometric evidences^{5,6,8,9}. The type section of these intertrappeans is at Takli, Nagpur. The beds have a similar stratigraphic arrangement throughout their exposures and are considered to lie near the Cretaceous/Tertiary (K/T) transition¹⁰. The reports of dinosaur remains from a number of intertrappeans¹¹⁻¹³ support a late Maastrichtian rather than a Tertiary age. This date is consistent with palaeomagnetic and radiometric studies of equivalent intertrappean horizons^{7,9}. The type section at Takli lies within the 29R magnetic zone encompassing the K/T boundary.

Deccanolestes hislopi gen. et sp. nov.

Holotype. Right upper molar M^1 (NKIM/10).

Paratype. Right upper molar M^3 (NKIM/11).

Referred material. Right lower premolar P_3 (NKIM/12).

Horizon and locality. Terminal Cretaceous intertrappean horizon, Naskal, Pargi Taluq, Rangareddi District, Andhra Pradesh, India.

Diagnosis. Upper molars and premolar small (M^1 , width (W) = 1.41 mm, length (L) = 0.9 mm; M^3 , W = 1.3 mm, L = 0.81 mm; P_3 , W = 0.9 mm, L = 0.41 mm). M^1 with transverse crown, triangular in occlusal view, broad styler shelf, deep ectoflexus with a slightly larger anterior part and somewhat smaller posterior part, stylocone (styler cusp B) well developed, and most prominent cusp in the styler region. Parastyle (styler cusp A) small, basal but distinct, metastyle weak, paracone higher than metacone with steeply sloping lingual walls and connate basally. Conules small, protocone moderately developed with an anteriorly and dorsally trending apex, cingula absent (Fig. 2, a-d; Fig. 3, d-f). M^3 with well developed parastylar area, wide styler shelf, shallow ectoflexus dividing the styler shelf into a much larger labially projecting anterior part and a smaller posterior part. Paracone twice as large as metacone, paraconule winged extending farther linguallly than the metaconule. Protocone is low and has slightly anteriorly shifted coronal geometry

Received 2 November 1987; accepted 26 January 1988.

- Fuoss, R. M. *J. Am. chem. Soc.* **80**, 5059-5061 (1968).
- Fournier, R. O. & Potter, R. W. II *Geochim. cosmochim. Acta* **46**, 1969-1974 (1982).
- Sommerfeld, R. A. *J. geophys. Res.* **72**, 4253-4257 (1967).
- Shettel, D. L. thesis, Pennsylvania State Univ. (1974).
- Novgorodov, P. G. *Geokhimiya* **10**, 1484-1489 (1975).
- Walther, J. V. & Orville, P. M. *Am. Miner.* **68**, 731-741 (1983).
- Crerar, D. A. & Anderson, G. M. *Chem. Geol.* **8**, 107-122 (1971).
- Marshall, W. L. & Warakowski, J. M. *Geochim. cosmochim. Acta* **44**, 915-924 (1980).
- Marshall, W. L. *Geochim. cosmochim. Acta* **44**, 925-931 (1980).
- Anderson, G. M. & Burham, C. W. *Am. J. Sci.* **265**, 12-27 (1967).
- Anderson, G. M. & Burham, C. W. *Am. J. Sci.* **283A**, 283-297 (1983).
- Dandurand, J. L. & Schott, J. J. *J. sol. Chem.* **16**, 237-256 (1987).
- Schott, J. & Dandurand, J. L. in *Chemical Transport in Metasomatic Processes* (ed. Helgeson, H. C.) 733-754 (Reidel, Dordrecht, 1987).
- Born von, M. Z. *Phys.* **1**, 45-48 (1920).
- Brønsted, J. N. *J. Am. chem. Soc.* **44**, 877-880 (1922).
- Kirkwood, J. G. *J. chem. Phys.* **7**, 911-919 (1939).
- Deul, R. thesis Univ. Karlsruhe (1984).
- Looyenga, H. *Physica* **31**, 401-404 (1965).
- Flowers, G. C. *Contr. Miner. Petrol.* **69**, 315-318 (1979).
- Pitzer, K. S. *Proc. natn. Acad. Sci. U.S.A.* **80**, 4575-4576 (1983).
- Bjerrum, N. J. *Kgl. Danske Videnskab. Selskab. Math-fys.* **7**, 1-48 (1926).
- Marshall, W. L. & Quist, A. S. *Proc. natn. Acad. Sci. U.S.A.* **58**, 901-903 (1967).
- Quist, A. S. & Marshall, W. L. *J. phys. Chem.* **72**, 1536-1544 (1968).
- Fournier, R. O. & Marshall, W. L. *Geochim. cosmochim. Acta* **47**, 587-596 (1983).
- Marshall, W. L. & Mesmer, R. E. *J. sol. Chem.* **10**, 121-127 (1981).
- Quist, A. S. & Marshall, W. L. *J. phys. Chem.* **72**, 684-703 (1968).
- Frantz, J. D. & Marshall, W. L. *Am. J. Sci.* **282**, 1666-1693 (1982).
- Frantz, J. D. & Marshall, W. L. *Am. J. Sci.* **284**, 651-667 (1984).
- Frantz, E. U. *Z. phys. Chem. Neue Folge* **8**, 107-126 (1956).
- Frantz, E. U. *Z. phys. Chem. Neue Folge* **8**, 192-206 (1956).

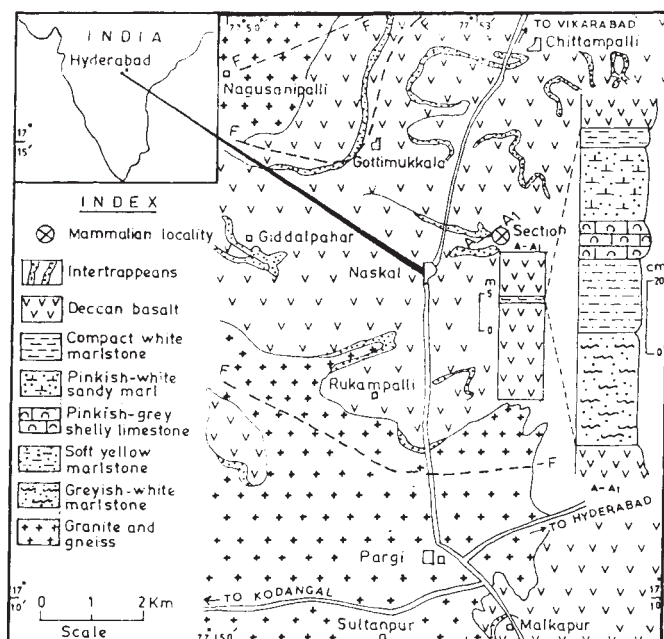


Fig. 1 Map showing the location of the mammal-bearing intertrappean horizon along with the measured stratigraphic section.

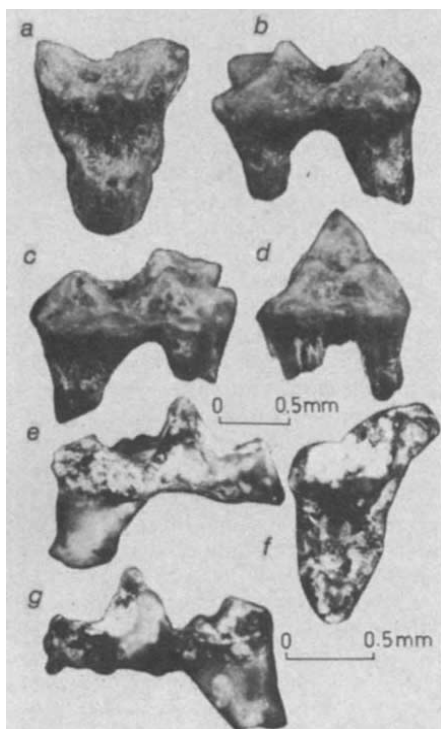


Fig. 2 *Deccanolestes hislopi* gen. et sp. nov. a-d right upper molar M^1 . a, occlusal view. b, posterior view. c, anterior view. d, labial view. e-g, right upper molar M^3 . e, anterior view. f, occlusal view. g, posterior view.

(Fig. 2, e-g; Fig. 3, a-c). P_3 with high lingually situated protoconid and a deep talonoid that separates the protoconid from the posterior cusps (Fig. 3, g-h). In several features the teeth may be considered primitive: a large W/L ratio, the development of a broad stylar shelf, large stylocone, paracone and metacone with connate, steeply sloping base lingually, only slightly differentiated labially, distinct but low protocone, small conules and absence of protoconular cingula. These features

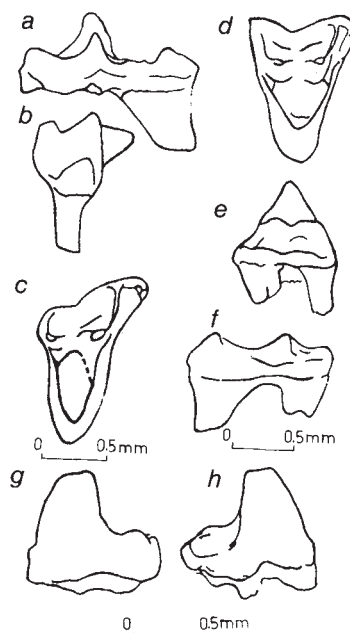


Fig. 3 *Deccanolestes hislopi* gen. et sp. nov. a-c, right upper molar M^3 . a, posterior view. b, lingual view. c, occlusal view. d-f, right upper molar M^1 . d, occlusal view. e, labial view. f, anterior view. g-h, right lower premolar P_3 . g, lingual view. h, labial view.

with noticeable similarities to corresponding teeth of *Cimolestes* and *Kennalestes* support a palaeoryctid affinity for the teeth.

This find demonstrates the wide distribution of primitive mammals with a tribosphenic pattern, both in Laurasia and Gondwanaland. Much of the non-mammalian fauna and flora of the Indian plate during the K/T transition also shows affinities to contemporary biota from other parts of the world¹⁰. For example, palaeoryctid mammals with affinities to late Cretaceous *Batodon* of North America have been reported from the early Palaeocene sediments of the Ouarzazate Basin, Morocco¹⁴ and were found in association with non-mammalian vertebrates remarkably similar to those recovered from the mammal-bearing Naskal intertrappeans. The dinosaurs of India have definite South American affinities¹⁵. Elements of the Deccan intertrappean flora have similarities to the late Cretaceous flora of South America¹⁶, Africa¹⁷ and Australia¹⁸. Otoliths from the Naskal intertrappeans show affinities to osteoglossids related to a living Australian taxa *Notopterus* and to a South American genus *Osteoglossum* (R. S. Rana, personal communication). Conversely, the ostracodes and charophytes from the same beds suggest close affinities to taxa from the late Cretaceous Nemegt Basin of Mongolia as suggested by the common presence of *Nemegtichara* (S. B. Bhatia, personal communication).

The lack of biotic endemism in the Indian plate has recently been interpreted to imply an India that was never far removed from its present position⁴. However, we feel that this hypothesis disregards the well-established palaeomagnetic and sea-floor spreading evidence which shows that India was in the southern tropical belt at about 29° S during the K/T transition. This latter evidence implies that any Laurasian taxa in the Indian plate must have been dispersed through a long-ranging dispersal corridor⁵. However, if one accepts the hypothesis of a stable India, it is equally difficult to explain the occurrence of Gondwanic elements. We feel that a solution to the dispersal of terrestrial biotas into the drifting Indian plate can only become possible when the geotectonic elements (microplates, island arcs) along the northern margin of Gondwanaland and western margin of the Indian plate have been precisely identified in space and time.

We thank Professors William A. Clemens, Jason A. Lillegraven, Zofia Kielan-Jaworowska, Jean-Jacques Jaeger and Emmanuel Gheerbrant for their help and the Council of Scientific and Industrial Research, New Delhi for financial support.

Received 7 December 1987; accepted 10 February 1988.

1. Clemens, W. A., Lillegraven, J. A. & Lindsay, E. H. in *Mesozoic Mammals* 7–58 (University of California Press, 1979).
2. Sige, B. *C.r. Acad. Sci.* **273**, 2479–2481 (1971).
3. De Muizon, C. *et al. Geobios* **16**, 747–753 (1983).
4. Chatterjee, S. & Hotton, N. *J. Southeast Asian Earth Sci.* **1**, 145–189 (1986).
5. Sahni, A. *et al. Bull. Soc. Geol. France* **24**, 1093–1099 (1982).
6. Gayet, M., Rage, J. C. & Rana, R. S. *Mem. Geol. Soc. France N.S.* **147**, 55–65 (1984).
7. Besse, J. *et al. Proc. 1st Int. Meet. IGCP Project 216: Global Biological Events in Earth History*, 365–370 (1986).
8. Prasad, G. V. R., Sahni, A. & Gupta, V. J. *Geoscience J.* **7**, 163–180 (1986).
9. Courtillot, V. *et al. Earth planet. Sci. Lett.* **80**, 361–374 (1986).
10. Sahni, A. *Science* **226**, 441–443 (1984).
11. Lydekker, R. *Rec. geol. Surv. India* **23**, 20–22 (1890).
12. Rao, B. R. J. & Yadagiri, P. *Mem. Geol. Soc. India* **3**, 287–291 (1981).
13. Sahni, A., Rana, R. S. & Prasad, G. V. R. *J. pal. Soc. India* **29**, 26–33 (1984).
14. Capetta, H. *et al. Geobios* **11**, 257–263 (1978).
15. Van Valen, L. & Sloan, R. E. *Nature* **270**, 193 (1977).
16. Prakash, U. *J. Paleont.* **34**, 1027–1040 (1960).
17. Srivastava, S. K. *Proc. Symp. on Cretaceous of India* (Birbal Sahni Institute of Palaeobotany, 1983), 141–157.
18. Bande, M. B., Mehrotra, R. C. & Prakash, U. *The Palaeobotanist* **35**, 1–12 (1986).

Female choice selects for male sexual tail ornaments in the monogamous swallow

Anders Pape Møller

Department of Zoology, Uppsala University, Box 561, S-751 22 Uppsala, Sweden

Male secondary sexual ornaments are widespread among polygynous animals¹ where they apparently arise through female choice and differential mating success of males^{2–5}. Darwin¹ and Fisher² suggested that female choice should also select for extravagant male ornaments in monogamous birds because female preference for ornamented males should result in earlier breeding which would increase male reproductive success by enhancing the quality and/or the quantity of offspring raised each year. In such cases female reproductive success should be positively related to body condition⁶ and a heritable component of the time of breeding should be affected by stabilizing selection⁷. Female preferences for ornamented males would be favoured by sexual selection because such choosy females would have a greater probability of bearing ornamented sons with an increased chance of pairing up early. I report here that in the monogamous swallow, *Hirundo rustica*, males with experimentally elongated tail ornaments obtain mates more quickly than males with shorter tails, and enjoy increased reproductive output in one breeding season. Such males are also preferred by females seeking extra-pair-bond copulations. Thus male sexual ornaments may also be maintained by female mate choice in monogamous species.

My experiments were performed at Kraghede, Denmark, at eight sites each with 1–17 breeding pairs of swallows⁸. The swallow is a small (~20 g body mass) insectivorous passerine that feeds on the wing and usually nests in small colonies. Population size in the study area varied markedly from 12 to 70 pairs during 1970–87. The number of breeding pairs per colony site ranged from 1 to 38 during this period. More than 90% of all pairs remained together during both first and second clutches in the same breeding season, whereas mate fidelity between years was low, primarily due to high overwinter mortality (A.P.M., manuscript in preparation). Sexual size dimorphism is slight⁹, except for the outermost tail feathers on each side which are 16% longer in the male (mean 105 mm, $N = 74$) than

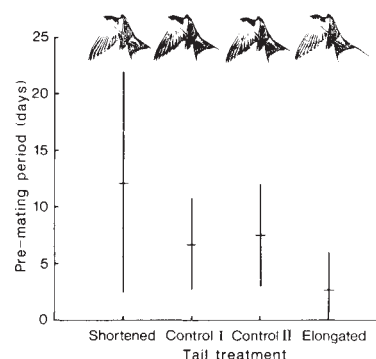


Fig. 1 Length of pre-mating period (days from arrival to pairing) for male swallows in relation to tail treatment. Sample sizes were 11 males in each case except the group with shortened tails ($N = 9$) in which two unmated males had to be excluded. Values are means (horizontal lines) $\pm$ s.d. (vertical lines). Length of pre-mating period differed among groups ($P < 0.01$, analysis of covariance using original tail length as a covariate); however, mean tail length did not differ between treatments before manipulation ($P > 0.10$, one-way analysis of variance), but did so after manipulation ($P < 0.001$). All groups differed in pairwise comparisons ($P < 0.05$ in each case, Mann-Whitney U -test), except the two control groups ($P > 0.10$).

in the female (mean 91 mm, $N = 63$). Tail length is significantly more variable in the male (coefficient of variation (c.v.) = 9.6%, $N = 74$) than in the female (c.v. = 6.1%, $N = 63$; $P < 0.01$, t -test) and than wing length and tarsus length of both sexes (c.v. = 2.3–4.8%; $P < 0.001$). Male swallows establish breeding territories of a few square metres, and attempt to attract a mate by tail displays performed either in the air or while perched⁹. Males were visited by from one to three females prospecting for a mate (mean (s.d.) = 1.17 (0.34), $N = 74$) before females made their choice. Male tail length was not related to territory size nor to his investment in nest building, mate guarding, or the feeding of nestlings (A.P.M., in preparation). Early-arriving, older males have longer tails than late-arriving, younger males (male tail length versus date of arrival: $r = -0.59$, $N = 74$, $P < 0.001$; male tail length versus age of male: $r = 0.51$, $N = 74$, $P < 0.001$).

To perform the experiments male swallows were caught at night while they were roosting within their territories. In all 44 colour-ringed males were randomly assigned to one of four groups. The first group (shortened) had a 20-mm long piece of feather cut from the middle of each of the two longest tail feathers beginning at ~2 cm from the base of the feather. The apical pieces were glued back on the original feather using cyanoacrylate superglue and liquid self-cure acrylic repair material as a catalyst. Hardening of the glue took less than 1 second. Mean tail length was reduced by 19.8% in this group to 85 mm. The second group of males (elongated) had 20-mm pieces added to the middle of their outermost tail feathers. These feathers were cut ~2 cm from their base and the new pieces glued in as above, resulting in an average increase in tail length of 16.5% to 127 mm. The third group of males (control group I) had their outermost tail feathers cut at ~2 cm from their base and glued back again to control for treatment effects. This resulted in an average 1% loss in tail length to 106 mm. The fourth group (control group II) were captured but their tails were not manipulated and had an average length of 106 mm. Manipulated tail lengths (76–138 mm) were approximately the same as the populations's natural range (84–132 mm; $N = 74$).

Male mating success was measured as the duration of the pre-mating period, that is, time from arrival of the male in the breeding colony until permanently paired with a female. On average, males with shortened tails took more than four times as long to acquire a mate than males with elongated tails (Fig. 1). Only two males remained unmated and both of these had their tails shortened. Females rejected tail-shortened males more

Fig. 2 *a*, Proportion of males having second clutches in relation to tail treatment. Sample sizes as in Fig. 1. The proportion of males having second clutches differed among groups ($P < 0.02$, χ^2 -test). Tail-shortened males less often achieved second clutches than controls ($P = 0.03$, Fisher exact probability test) and than males with elongated tails ($P = 0.003$). Other groups did not differ from each other ($P > 0.10$ in each case). *b*, Annual reproductive success (number of fledglings) of male swallows in relation to tail treatment. Sample sizes as in Fig. 1. Values given are mean (horizontal lines) $\pm$ s.d. (vertical lines). The reproductive success differed among groups ($P < 0.001$, analysis of covariance using original tail length as a covariate). All groups differed in pairwise comparisons ($P < 0.05$ in each case, Mann-Whitney *U*-test), except the two control groups ($P > 0.10$).

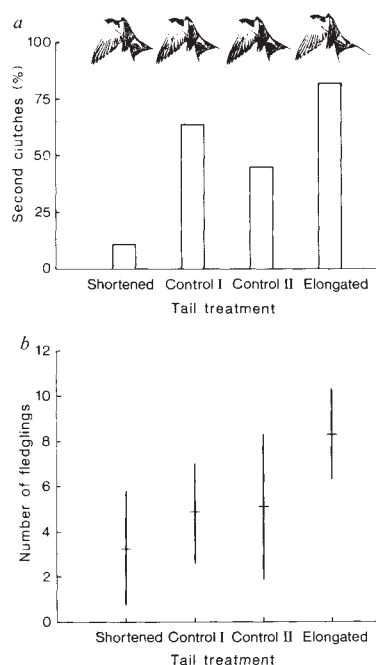


Table 1 Effect of tail treatment of male swallows on copulation behaviour

Copulation behaviour	Tail treatment				<i>P</i>
	Shortened	Control I	Control II	Elongated	
Male intra-pair copulation attempt rate	5.9 (11.3)	4.8 (10.3)	4.5 (9.8)	3.5 (5.9)	>0.10
Male intra-pair copulation rate	1.3 (2.6)	1.1 (2.4)	1.0 (2.1)	0.8 (1.3)	>0.10
Male extra-pair copulation attempt rate	0.10 (0.16)	0.12 (0.13)	0.11 (0.09)	0.14 (0.11)	>0.10
Male extra-pair copulation rate	0	0	0	0.040 (0.048)	<0.001
Female extra-pair copulation rate	0.030 (0.036)	0.004 (0.014)	0.005 (0.017)	0	<0.01

Values are means (1 s.d.) for 1-h observations daily during the period from five days before start of egg-laying until the day when the penultimate egg was laid (intra-pair copulation behaviour, female extra-pair copulation behaviour), or during the period from five days before start of egg-laying until the fifth day of the incubation period (male extra-pair copulation behaviour). *P* values calculated from an analysis of covariance using original tail length as a covariate.

female mates of control and especially of short-tailed males. These results suggest that female swallows use male sexual ornamentation as a cue in their choice of extra-pair copulation partners, thus providing a second selective force favouring such ornamentation.

Although males obviously gained by pairing up as soon as possible after arrival, it is not obvious why female swallows should prefer males with long tail ornaments. Females may prefer such males if the size of their sexual ornament reflects either their overall genotypic or phenotypic quality¹³⁻¹⁶. Such quality might be reflected in resistance to parasites¹⁷, nutritional state, contribution to parental care or simply overall viability¹⁵. Long tail feathers could therefore be conditional characters¹⁸ and need not be heritable^{18,19} to provide some advantage to females selecting such males.

Differences in ornament size could also reflect differential combat ability and thus could possibly have resulted from intrasexual selection^{11,15}. But all male swallows establish a territory, irrespective of their tail size, and because fledging success is high, there appears to be little variation in territory quality. Interactions between males during the breeding season are primarily territory intrusions during the fertile period of a female²⁰. Because such intrusions did not result in territory take-overs or different rates of extra-pair copulation attempts between groups and territory quality did not vary, sexual ornaments are not likely to have evolved due to intrasexual selection.

I conclude that tail ornaments in the monogamous swallow have evolved as a result of female choice. Because females prefer males carrying a slightly enlarged sexual ornament, such males will pair up earlier^{3,21}, have an enhanced probability of siring two broods and thus produce more offspring within one breeding season. These results are consistent with the Darwin-Fisher-O'Donald idea^{1,2,6,21} that males with longer tail feathers will be able to pair up earlier with females having superior non-heritable quality. It is highly unlikely that more attractive males get mates with high heritable viability, although it is possible that males with long tail feathers do get mates with superior quality that equates with double clutching ability. The widespread occurrence of male secondary sexual ornaments in monogamous birds¹ may similarly have arisen and be maintained by intersexual selection.

T. R. Birkhead, L. Gustafsson, J. Höglund, A. Lundberg, R. D. Montgomerie, A. Pomiankowski, S. Ulfstrand and J. Waage provided constructive criticism. A. Ulfstrand made the drawings. M. Andersson helped with information on feather glue. The study was supported by the Swedish NSRC.

Received 18 December 1987; accepted 29 February 1988.

1. Darwin, C. *The Descent of Man, and Selection in Relation to Sex* (Murray, London, 1871).
2. Fisher, R. A. *The Genetical Theory of Natural Selection*. 2nd Ed (Dover, New York, 1958).

frequently (mean (s.d.) = 0.73 (0.65) female rejections per male, $N = 11$) than males in any other group (0(0) female rejections per male, $N = 33$; $P < 0.001$, Mann-Whitney *U*-test).

A short pre-mating period should allow earlier egg laying, and males with elongated tails did have two clutches significantly more often than males with shortened tails (Fig. 2*a*). Total reproductive success of males with elongated tails (measured as the total number of fledglings in one breeding season) was more than twice as high as in males with shortened tails (Fig. 2*b*). Other measures of reproductive success (clutch size, brood size at hatching, brood size at fledging, and both mean tarsus length (=adult size) and mean body mass of nestlings aged 15 days) did not differ among treatments, either for first or second clutches ($P > 0.10$, Kruskal-Wallis one-way analysis of variance). Thus survival prospects for fledglings should not have differed among treatments as survival is related to body size and body weight¹⁰.

Although these results seem clear, female swallows may have chosen males on the basis of their display frequency¹¹⁻¹², arrival date, or territory quality¹¹ irrespective of tail length, even though males were randomly assigned to treatment groups. To test this, male display rates were measured during one-hour observation periods every day during the pre-mating period. But song rate per hour and tail display rate per hour did not differ significantly among experimental groups ($P > 0.10$, Kruskal-Wallis one-way analysis of variance). Male display rate thus was not affecting female choice. Mean arrival date of males also did not differ significantly among treatments ($P > 0.10$, analysis of covariance with tail length as a covariate).

It has been suggested that sexual selection may occur in apparently monogamous systems if females engage in extra-pair-bond copulations with favoured males²². Female swallows do not actively solicit copulations, but respond to male approachers, either accepting copulation attempts by remaining perched or rejecting them by flying away or behaving aggressively⁹. The rate of accepted and rejected intra-pair-bond copulation attempts did not differ significantly between groups of males (Table 1). Male swallows attempted extra-pair copulations at a similar rate. But extra-pair copulation rates were significantly higher among males with elongated tails as compared to other males. These extra-pair-bond copulations were directed at

3. O'Donald, P. *Genetic Models of Sexual Selection* (Cambridge University Press, 1980).
4. Lande, R. *Proc. natn. Acad. Sci. U.S.A.* **79**, 3721-3725 (1981).
5. Andersson, M. *Nature* **299**, 818-820 (1982).
6. Turner, A. K. *J. Anim. Ecol.* **51**, 29-46 (1982).
7. van Noordwijk, A. J., van Balen, J. H. & Scharloo, W. *Oecologia* **49**, 158-166 (1981).
8. Möller, A. P. *Anim. Behav.* **35**, 819-832 (1987).
9. von Vietinghoff-Riesch, A. *Die Rauchschnalbe* (Duncker and Humblot, Berlin, 1955).
10. Perrins, C. M. *J. Anim. Ecol.* **34**, 601-647 (1965).
11. Searcy, W. M. *A. Rev. Ecol. Syst.* **13**, 57-85 (1982).
12. Searcy, W. M. & Andersson, M. *A. Rev. Ecol. Syst.* **17**, 307-333 (1986).
13. Williams, G. C. *Adaptation and Natural Selection* (Princeton University Press, 1966).
14. Trivers, R. L. in *Sexual Selection and the Descent of Man, 1871-1971* (ed. Campbell, B.) 136-179 (Aldine, Chicago, 1972).
15. Andersson, M. *Biol. J. Linn. Soc.* **17**, 375-393 (1982).
16. Kodric-Brown, A. & Brown, J. H. *Am. Nat.* **124**, 303-323 (1984).
17. Hamilton, W. D. & Zuk, M. *Science* **218**, 384-387 (1982).
18. Andersson, M. *Evolution* **40**, 804-816 (1986).
19. Dominey, W. J. *J. theor. Biol.* **101**, 495-502 (1983).
20. Möller, A. P. *Behav. Ecol. Sociobiol.* **17**, 401-408 (1985).
21. O'Donald, P. *Nature* **237**, 349-351 (1972).

Selective responses of visual cortical cells do not depend on shunting inhibition

Rodney J. Douglas, Kevan A. C. Martin
& David Whitteridge

MRC Anatomical Neuropharmacology Unit,
Department of Pharmacology, South Parks Road,
Oxford OX1 3QT, UK

Theoretical analyses of the electrical behaviour of the highly branched processes of nerve cells has focused attention on the possibility that single cells perform complex logical operations rather than simply summing their synaptic inputs¹⁻³. In particular, it has been suggested that the orientation and direction selectivity of cells in the visual cortex result from the action of a nonlinear 'shunting' inhibition that emulates an AND-NOT logical operation⁴. The characteristic biophysical feature of this proposed inhibitory mechanism is that it evokes a large and relatively sustained increase in the conductance of the neuronal membrane while leaving the membrane potential unaffected. This shunting mechanism contrasts with linear 'summative' inhibition in which conductance changes are less prominent, and inhibition is achieved by hyperpolarization of the membrane potential¹. In a direct experimental test of the hypothesis that the selectivity of visual cortical neurons depends on shunting inhibition we found no evidence for the large conductance changes predicted by the theory.

To test for the presence of shunting inhibition, we measured the input conductance and membrane voltage of single neurons during the passage of a bar or edge stimulus at various orientations and directions of motion across the receptive field. The results reported here were obtained from 195 stimulus trials in which measurements were taken from eleven neurons (eight simple cells, three complex cells, layers 2-6) in six cats. The results for a single neuron, which exhibited many of the functional properties characteristic of others tested, are presented in Figs 1-3. Extracellular recording using dark and light bars (Fig. 1a, b) confirmed that this neuron had an S-type, or simple receptive field with separate 'on' and 'off' subfields. This neuron was identified by HRP (horseradish peroxidase) filling as a layer 6 pyramidal cell. Figure 1c, d indicates that a light bar moving in the preferred direction induces a small hyperpolarization of about 5 mV (arrowed) when leaving the 'on' subfield. Inhibition of discharge is not associated with the large sustained conductance increase expected of shunting inhibition (Fig. 1d, e).

Extracellular data and theoretical analyses have been used to support the hypothesis that the shunting mechanism is responsible for the direction and orientation sensitivity of cortical neurons⁴⁻⁷. This mechanism would permit the neuron to respond to a stimulus of preferred orientation and direction of movement, but would inhibit a response to a non-optimal stimulus. Comparison of the action potential discharge in the intracellular

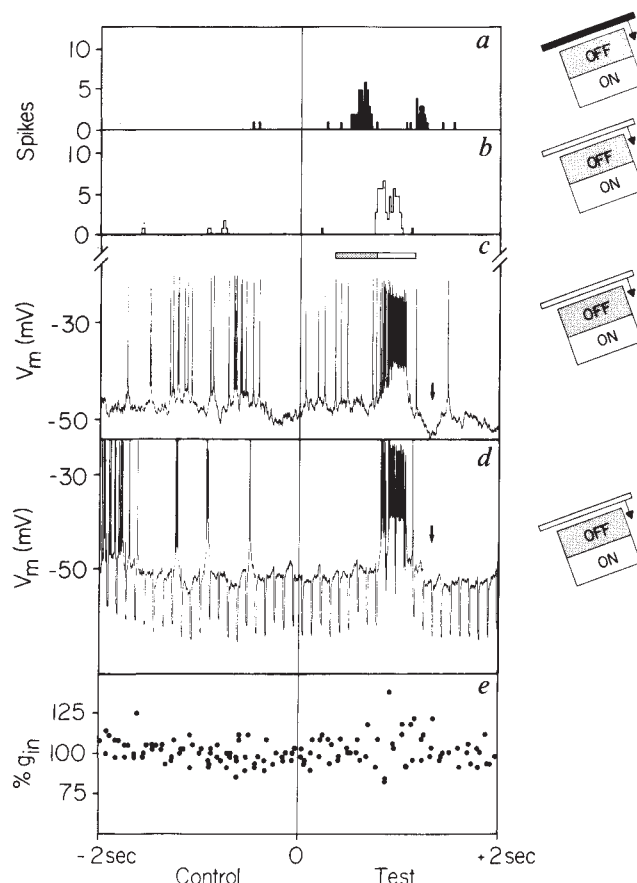


Fig. 1 The responses of a simple cell to a bar moving across its receptive field in the preferred orientation and direction. Recordings are from a layer 6 pyramidal neuron. The stimulus was a 0.29×11.7 degree bar moving at 2.9 degree s^{-1} . Contrast, orientation and the direction of motion of bar with respect to the receptive field are indicated on schematics (not to scale). The positions of 'on' and 'off' subfields are indicated over the intracellular recordings. Current pulse capacitance artefacts have been removed. *a, b*, Extracellular averaged response to dark and light bars (five trials). *c, d*, Intracellular recordings show sustained depolarization and action potential discharge as the light bar passed through 'on' subfield (compare *b*), and hyperpolarization (arrowed) as the bar left the 'on' subfield. *d*, Constant current pulses (-0.1 nA) were used to measure input conductance (g_{in}). *e*, Conductance measurements from three trials staggered in time, were normalized against the mean conductance of their control periods and expressed as $\% g_{in}$. There was no evidence of the sustained large conductance change expected of shunting inhibition. Receptive field description: light edge, and light bar (*b*), moving in the preferred direction elicited a single 'on' subfield. Dark edge moving in preferred direction elicited one discharge peak, but dark bar leaving 'on' subfield elicited two peaks (*a*). Thus the minor peak was not due to second 'off' subfield, but to dark bar leaving 'on' subfield. Light bar moving out of 'on' subfield inhibited the discharge (*c, d* arrowed). Inhibition induced when light bar moved through the 'off' subfield was partly counteracted by the excitation induced by the (dark) trailing edge of the light bar⁸.

Methods. Neurons were recorded from the postlateral gyrus of the striate visual cortex (area 17) of anaesthetized, paralysed cats²¹, while continuously monitoring vital signs. Glass micropipettes were filled with 2 M potassium citrate, or a 4% buffered solution of HRP in 0.2 M KCl. Receptive fields of the cortical neurons were first plotted in detail by hand. Extracellular responses to computer-controlled visual stimuli were quantified by peristimulus time histograms. Thereafter, intracellular recordings were made during presentation of the same computer-controlled stimuli. Input conductance measurements were derived from the increment in membrane voltage evoked by hyperpolarizing current pulses (20-30 ms, 0.1-0.5 nA). Where possible, HRP was injected intracellularly following data collection to enable morphological identification.

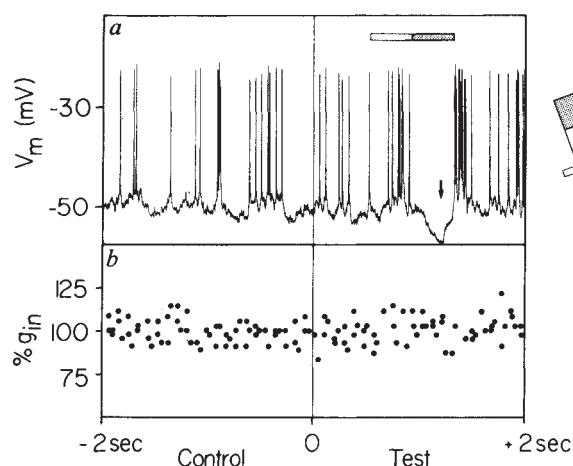


Fig. 2 A light bar moves across the simple cell receptive field in the non-preferred direction. *a*, Intracellular recording shows reduced response compared to that of preferred direction (Fig. 1c), and hyperpolarization (arrowed) as light bar moved through 'off' subfield. 'on' and 'off' zones over recording are reversed with respect to Fig. 1c because bar moved across receptive field in opposite direction. *b*, Conductance measurements show no evidence of sustained large changes as bar moved across receptive field. Thus the diminished response of the neuron to non-preferred direction was not due to shunting inhibition.

records of Fig. 1c and Fig. 2a indicates that this neuron had a direction preference for the light bar. This preference could be predicted from the arrangement of the 'on' and 'off' subfields⁸. There was no evidence of the sustained large increase in the input conductance expected of shunting inhibition as the light bar moved in the non-preferred direction (Fig. 2b). However there was a clear hyperpolarization and inhibition of the discharge (arrowed) as the bar passed through the 'off' subfield (Fig. 2a). A similar absence of a large conductance change was found when the stimulus was presented at 20 degrees, or 90 degrees, (Fig. 3a, b) with respect to the preferred orientation.

For the shunting conductance to be effective, it must be sustained for the duration of the flow of the excitatory current.⁹ For both direction and orientation sensitivity, the shunt would have to be sustained for at least the time the stimulus took to traverse the relevant dimension of the receptive field. The sampling density of our current pulses was high enough to detect conductance changes of these durations. In addition, a shunting conductance would need to be large (10–100 nS)^{2,4}. A 10 nS conductance at the soma of the neuron illustrated here would have produced a change of nearly 100% in its input conductance. Our methods could detect changes in input conductance of greater than ~15%. Large sustained conductance changes were not found, irrespective of the receptive field type of the neuron (simple or complex), the contrast of the stimulus, or the receptive field property being examined (subfield antagonism of simple cells, orientation tuning, or direction preference).

One explanation of our results is that the neurons we recorded from were not being inhibited by postsynaptic mechanisms. Instead we may have detected a withdrawal of excitation due to inhibition at an earlier stage in the circuit. This could certainly be an explanation for the trailing hyperpolarization arrowed in Fig. 1c. There is, however, strong pharmacological^{10–13} and immunocytochemical^{14,15} evidence for postsynaptic GABA-ergic inhibitory mechanisms and it would be surprising if they were not activated under the stimulus configurations we have used.

Another possible explanation is that the shunting synapses were distant from the soma and were thus masked by the impedance properties of the intervening dendritic tree. An extreme example would have the shunting synapse located on

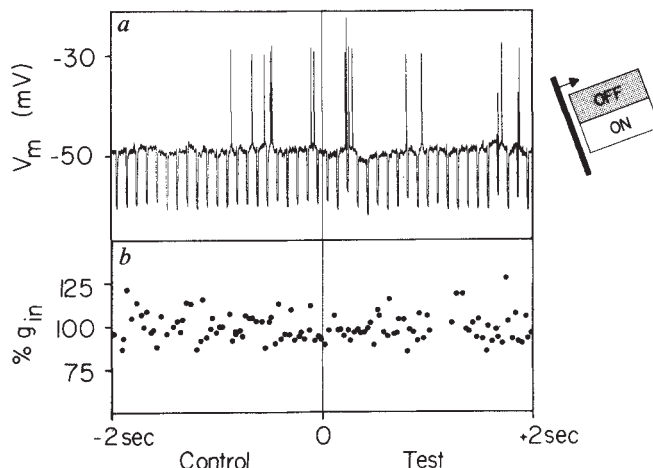


Fig. 3 A dark bar oriented at 90 degrees to the optimal moves across the simple cell receptive field. *a*, Intracellular recording shows no response by comparison with that of preferred orientation (Figs 1c and 2a). *b*, Conductance measurements show no evidence of sustained large change as bar moved across receptive field. Thus failure of the neuron to respond to the cross-oriented stimulus was not due to shunting inhibition.

a spine head, so providing very specific inhibition of an excitatory synapse located on the same spine head. In this case the conductance change induced by both synapses would be masked from the soma by the high axial resistance of the spine neck. Even if shunting inhibition does occur on spines, its effect may be small; the main excitatory input to pyramidal cells arrives on spines^{16–19}, but only 7% of spines have an additional Gray's type 2 'inhibitory' synapse²⁰. The HRP-filled neurons in this and previous studies²¹ show that in most cases the recording electrode impaled the soma or the proximal dendrites. It is in these same regions that anatomical and immunocytochemical studies have revealed the major concentration of Gray's type 2, GABA-ergic synapses^{14–17,22,23}. Thus most inhibition should occur at the perisomatic region.

We are left with the conclusion that cortical inhibition, when it occurs, operates by small and/or transient conductance changes beneath the sensitivity of our sampling methods. One would predict that to produce an effective inhibition with such conductance changes, the reversal potential of the ion(s) concerned must be considerably more negative than the threshold for generating action potentials. However, the inhibition of response that we see is usually associated with quite small hyperpolarizations of the membrane and in some trials the membrane potential does not even fall below the 'resting' level, despite inhibition of discharge, as previously noted by Creutzfeldt and Ito²⁴. Thus the degree of hyperpolarizing inhibition may well be calibrated so as to reduce or cut off the firing of the cell, without driving the membrane potential to such negative values that the cell becomes insensitive to additional excitation.

We thank John Anderson for technical assistance, and the E. P. Abrahams Trust for support. R.J.D. acknowledges the support of the Guarantors of Brain, and the SA MRC.

Received 29 January 1988; accepted 4 March 1988.

1. Blomfield, S. *Brain Res.* **69**, 115–124 (1974).
2. Koch, C., Poggio, T. & Torre, V. *Phil. Trans. R. Soc. B* **298**, 227–264 (1982).
3. Swindale, N. V. *Nature* **303**, 570–571 (1983).
4. Koch, C. & Poggio, T. in *Models of the Visual Cortex* (eds Rose, D. & Dobson, V. G.) 408–419 (Wiley, Chichester, 1985).
5. Rose, D. *Expl. Brain Res.* **28**, 221–223 (1977).
6. Dean, A. F., Hess, R. F. & Tolhurst, D. J. *J. Physiol.* **236**, 84P (1980).
7. Emerson, R. C., Citron, M. C. & Felleman, D. J. in *Models of the Visual Cortex* (eds Rose, D. & Dobson, V. G.) 420–431 (Wiley, Chichester, 1985).
8. Hubel, D. H. & Wiesel, T. N. *J. Physiol.* **160**, 106–154 (1962).
9. Torre, V. & Poggio, T. *Proc. R. Soc. B* **202**, 409–416 (1978).
10. Rose, D. & Blakemore, C. *Nature* **249**, 375–377 (1974).

11. Sillito, A. M. *J. Physiol.* **250**, 305-322 (1975).
12. Tsumoto, T., Eckhart, W. & Creutzfeldt, O. D. *Expl. Brain Res.* **34**, 351-363 (1979).
13. Vidyasagar, T. R. & Heide, W. *Neuroscience* **17**, 49-55 (1986).
14. Ribak, C. E. *J. Neurocytol.* **7**, 461-478 (1978).
15. Freund, T. F., Martin, K. A. C., Smith, A. D. & Somogyi, P. *J. comp. Neurol.* **221**, 263-278 (1983).
16. Levay, S. *J. comp. Neurol.* **150**, 53-86 (1973).
17. Peters, A. in *Synaptic Function* (eds Edelman, G. M., Gall, W. E. & Cowan, W. M.) 373-397 (Wiley, New York, 1987).
18. Collonier, M. *Brain Res.* **9**, 268-287 (1968).
19. Szentagothai, J. in *Handbook of Sensory Physiology*. Vol. 7. *Central Visual Information* (ed. Jung, R.) 269-324 (Springer, Berlin, 1973).
20. Beaulieu, C. & Colonnier, M. *J. comp. Neurol.* **231**, 180-189 (1985).
21. Martin, K. A. C. & Whitteridge, D. *J. Physiol.* **353**, 463-504 (1984).
22. Martin, K. A. C., Somogyi, P. & Whitteridge, D. *Expl. Brain Res.* **50**, 193-200 (1983).
23. White, E. L. & Rock, M. P. *J. Neurocytol.* **9**, 615-636 (1980).
24. Creutzfeldt, O. & Ito, M. *Expl. Brain Res.* **6**, 324-352 (1968).

Type I phosphatidylinositol kinase makes a novel inositol phospholipid, phosphatidylinositol-3-phosphate

Malcolm Whitman*, C. Peter Downes†, Marilyn Keeler, Tracy Keller & Lewis Cantley

Department of Physiology, Tufts University School of Medicine, Boston, Massachusetts 02111, USA

† Smith Kline and French Research Ltd, The Frythe, Welwyn, Herts AL6 9AR, UK

The generation of second messengers from the hydrolysis of phosphatidylinositol-4,5-bisphosphate (PtdInsP₂) by phosphoinositidase C has been implicated in the mediation of cellular responses to a variety of growth factors and oncogene products¹⁻⁴. The first step in the production of PtdInsP₂ from phosphatidylinositol (PtdIns) is catalysed by PtdIns kinase. A PtdIns kinase activity has been found to associate specifically with several oncogene products, as well as with the platelet-derived growth factor (PDGF) receptor⁵⁻⁸. We have previously identified two biochemically distinct PtdIns kinases in fibroblasts, and have found that only one of these, designated type I, specifically associates with activated tyrosine kinases⁷. We have now characterized the site on the inositol ring phosphorylated by type I PtdIns kinase, and find that this kinase specifically phosphorylates the D-3 ring position to generate a novel phospholipid, phosphatidylinositol-3-phosphate (PtdIns(3)P). In contrast, the main PtdIns kinase in fibroblasts, designated type II, specifically phosphorylates the D-4 position to produce phosphatidylinositol-4-phosphate (PtdIns(4)P), previously considered to be the only form of PtdInsP (ref. 9). We have also tentatively identified PtdIns(3)P as a minor component of total PtdInsP in intact fibroblasts. We propose that type I PtdIns kinase is responsible for the generation of PtdIns(3)P in intact cells, and that this novel phosphoinositide could be important in the transduction of mitogenic and oncogenic signals.

Type I PtdIns kinase was found to catalyse the phosphorylation of PtdIns to a product that at first appeared to comigrate with PtdIns(4)P in one and two-dimensional thin-layer chromatography (TLC) systems^{5,7}, but further comparison showed the type I product migrated slightly more slowly (Fig. 1). A mixture of the two reaction products moved as a broad spot coincident with the PtdIns(4)P standard. Reverse-phase high-pressure (liquid chromatography (HPLC) of the type I and type II PtdIns phosphates shows that they have a similar distribution of fatty acyl side chains (M.W., G. Patten and M.K., unpublished observations). We have therefore investigated whether there is a difference in the head group of the two products. The PtdIns phosphates were synthesized using both enzymes and [³²P]ATP as phosphate donor, and after deacyla-

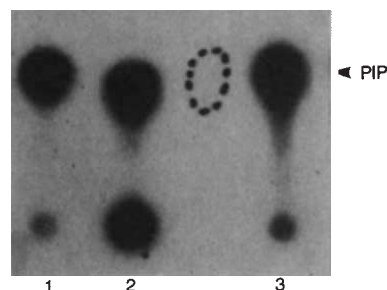


Fig. 1 Migration of PtdIns kinase products during TLC. Type I PtdIns kinase and type II PtdIns kinase were used to phosphorylate PtdIns with [³²P]ATP and the reaction products were extracted and separated by TLC. Lane 1: product of type II PtdIns kinase; lane 2: product of type I PtdIns kinase; lane 3: mixture of type I and type II PtdIns kinase products. dotted line indicates migration of cold PtdIns(4)P standard.

Methods. Type I PtdIns kinase was prepared by immunoprecipitation of middle T/pp60^{c-src} from middle T-transformed NIH 3T3 fibroblasts as described previously^{3,5}. Immunoprecipitates were washed⁵ and assayed in 0.2 mg ml⁻¹ sonicated detergent-free PtdIns (Avanti) with 10 μCi [³²P]ATP, 0.02 mM ATP, 10 mM MgCl₂, 0.1 M NaCl, 20 mM Tris, pH 7.6 in a final volume of 50 μl for 5 min. Type II PtdIns kinase was prepared by precipitation with Sepharose CL-4B from polyoma-transformed fibroblasts⁵. Precipitates were assayed as above except that 0.5% NP40 detergent was included and ATP was at 0.05 mM. After extraction, reactions were separated by TLC in CHCl₃:MeOH:2.5 M NH₄OH (9:7:2) and products were visualized by autoradiography; the PtdIns(4)P standard was visualized by staining with iodine vapour.

tion the resulting glycerophosphoinositol phosphates (GroPInsPs) were separated in an anion-exchange HPLC system optimized for resolution of GroPInsP isomers (Fig. 2). The deacylated type I product eluted about one minute earlier than an internal standard of [³H]GroPInsP prepared from turkey erythrocyte phospholipids, and presumed to be GroPIns(4)P (Fig. 2). Essentially all the PtdIns³²P produced by the type II kinase comigrated with the [³H]PtdInsP standard on deacylation and HPLC whereas the single earlier peak of the deacylated PtdIns³²P produced by the type I kinase preparation separated completely from the standard. These results establish a structural difference between the two types of phosphorylated PtdIns, and indicate that it resides in the site of phosphorylation on the inositol head group. Each kinase was therefore quite specific for the characteristic head group structure generated.

To determine the different sites phosphorylated on the inositol ring by the type I and type II PtdIns kinases, PtdIns was labelled in the inositol ring with ³H and phosphorylated by each enzyme using unlabelled ATP. The resulting [³H]PtdIns phosphates were then TLC-purified and deacylated, and the glycerol moiety removed by mild periodate oxidation¹⁰. The resulting inositol bisphosphate (InsP₂) isomers were then identified according to the scheme depicted in Fig. 3. This involves periodate oxidation of the inositol ring, reduction with borohydride, and dephosphorylation with alkaline phosphatase to yield polyols characteristic of the different InsP₂ isomers^{11,12}. These polyols can be separated (with the exception of enantiomeric pairs) by HPLC on a Brownlee polybore column, as shown in Fig. 3a. After this procedure, PtdIns(4)P yields two polyols with the chromatographic properties of altritol and iditol. These are the polyols expected from the partial cleavage of inositol-1,4-bisphosphate (Ins(1,4)P₂) by periodate. We obtained the same result using [³H]InsP₂ derived from the type II PtdIns kinase product; in each case, the ratio of [³H]altritol: [³H]iditol was 0.29. This confirms that the type II PtdIns kinase converts PtdIns to PtdIns(4)P.

Preliminary experiments using the [³H]polyol obtained from

* Present address: Department of Biochemistry and Molecular Biology, Harvard University, 7 Divinity Avenue, Cambridge, Massachusetts 02138, USA

Fig. 2 Anion-exchange HPLC deacylated type I PtdIns kinase product. *a*, ^{32}P -labelled type I. *b*, Migration of a co-injected ^3H -labelled internal standard from turkey erythrocytes labelled with $[^3\text{H}]\text{inositol}^{16}$; $[^3\text{H}]\text{GroPIns4P}$ elutes at 39 min, type I ^{32}P GroPInsP elutes slightly earlier than the standard, at ~ 38 min. DPS, disintegrations per second.

Methods. ^{32}P -labelled type I PtdIns kinase product was prepared and TLC-purified as described in Fig. 1. The region of the TLC plate containing the PtdInsP was cut out and incubated at 53°C for 50 min in 25% methylamine/MeOH/butanol. The deacylated lipid mixture was then dried *in vacuo* overnight, redissolved in 2 ml water and extracted twice with butanol/petroleum ether/ethyl formate (20:4:1). The aqueous phase was dried *in vacuo* and redissolved in 0.5 ml water for HPLC. The sample was spiked with a standard mixture containing $[^3\text{H}]\text{GroPIns}$, $[^3\text{H}]\text{GroPInsP}$ and $[^3\text{H}]\text{GroPInsP}_2$, prepared using a lipid extract from $[^3\text{H}]\text{inositol}$ -labelled turkey erythrocytes¹⁶ deacylated exactly as described above. Anion-exchange HPLC was carried out on a Whatman 5 Partisphere SAX column, the sample loaded in water, the column washed for 10 min with water and eluted with $0.42\text{ M }(\text{NH}_4)_2\text{HPO}_4$, pH 3.8, at 1 ml min^{-1} . The linear gradient used was $0\text{--}0.252\text{ M}$ over 60 min, followed by $0.252\text{--}0.42\text{ M}$ over 10 min.

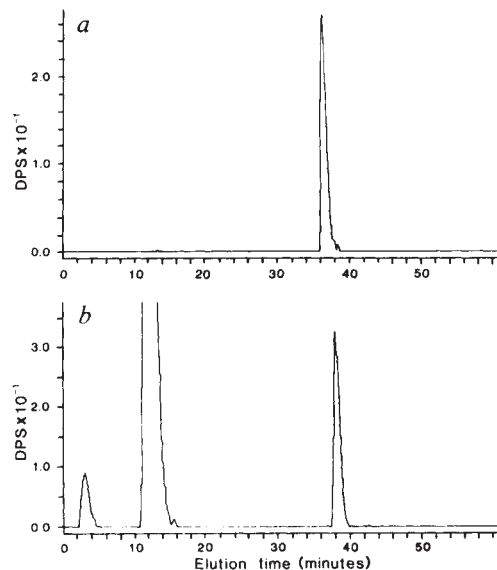
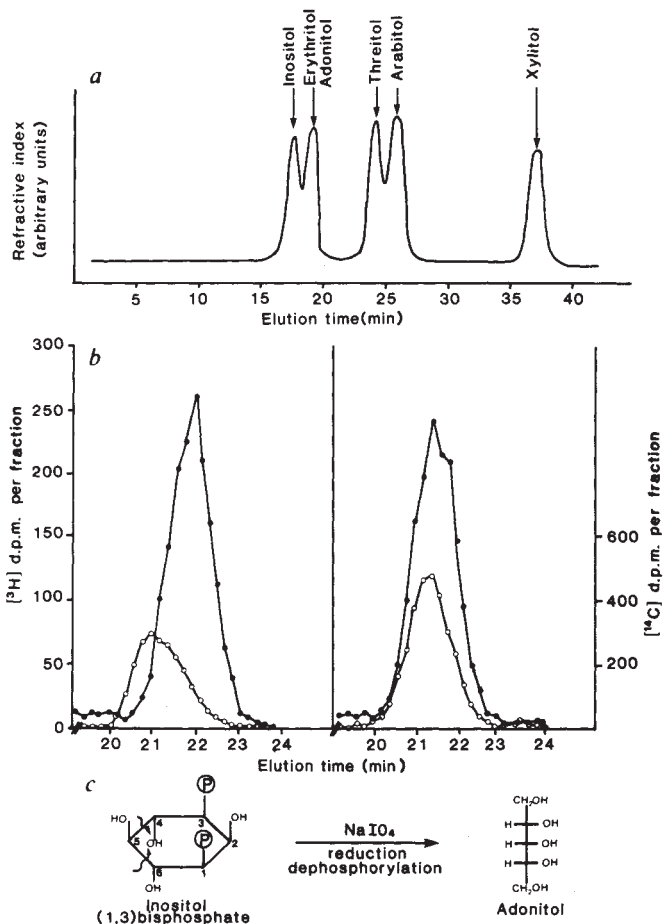


Fig. 3 Elucidation of the head group structure of type I PtdInsP. Panels *a*–*c* depict the strategy for the identification of InsP₂ isomers. Periodate oxidation of inositol bisphosphates, followed by reduction and dephosphorylation, yields a specific polyol that can be identified chromatographically. Of the 12 InsP₂ isomers not completely oxidized by periodate to yield CO₂ and P_i (so therefore not phosphorylated at opposite ends of the inositol ring), four should give threitol, three should yield xylitol, two should give erythritol and two arabitol. Only the optically inactive *meso*-compound, Ins(1,3)P₂, should give adonitol. *a*, HPLC separation of the five polyols and of myo-inositol, which is detectable when the periodate oxidation step is incomplete. *b*, Identification of $[^3\text{H}]$ polyol derived from type I PtdInsP as adonitol. The elution of two similar samples of the $[^3\text{H}]$ polyol is shown in the left and right panels (○), with the internal standard ^{14}C -labelled erythritol in the left panel (●), and ^{14}C -labelled adonitol in the right (●). *c*, Structure and derivation of adonitol from Ins(1,3)P₂.

Methods. $[^3\text{H}]\text{inositol}$ -labelled type I PtdInsP was prepared as described in Fig. 2, except that $[^3\text{H}]\text{PtdIns}$ (NEN) ($5\mu\text{g ml}^{-1}$, 4 Ci mmol^{-1}) and unlabelled ATP ($200\mu\text{M}$) were used as substrates. Fatty acids were removed from TLC-purified ^3H -labelled type I PtdInsP by transacylation with methylamine as described¹⁸. Glycerol was cleaved from the $[^3\text{H}]\text{GroPInsP}$ sample by mild treatment with sodium periodate¹⁰. Remaining aldehyde was removed by 1,1-dimethylhydrazine ($1.875\text{ ml } 1\% \text{ w/v}$) and incubation at 25°C for 4 h. The sample was desalted on a 3 ml column of Bio-Rad AG-50 ($200\text{--}400$ mesh), dried *in vacuo*, and the resulting $[^3\text{H}]\text{Type I InsP}_2$ purified by HPLC on a Whatman Partisphere FAX column eluted a linear gradient over 30 min, from $0\text{--}0.8\text{ M}$ triethylammonium formate (pH 3.8). $[^3\text{H}]\text{InsP}_2$ eluted about 20 min after the start of the gradient. The sample was dried and triethylammonium formate removed *in vacuo*. Identification of inositol phosphate isomers was as described¹¹, and those from ^3H -labelled type I InsP₂ were identified as described previously for inositol tetrakisphosphates¹⁹, except that the incubation with 100 mM NaIO_4 was at pH 4.5. The resulting ^3H -labelled polyol was characterized by chromatography on a Brownlee polybore carbohydrate column¹⁹, except that the column was equilibrated at 25°C rather than at 90°C . Standard polyols were from Sigma (myo-inositol); Aldrich (arabitol, erythritol and xylitol); and Lancaster Synthesis (D/L threitol); $[^{14}\text{C}]\text{erythritol}$ was supplied by Amersham and $[^{14}\text{C}]\text{adonitol}$ was prepared by reduction of $[^{14}\text{C}]\text{ribose}$ (Amersham) using NaBH_4 (ref. 19). Unlabelled standards ($25\mu\text{g}$) were injected using a $20\text{-}\mu\text{l}$ loop and detected using a differential refractometer (Waters, model 410). For separation and quantitation of ^{14}C -labelled standards and $[^3\text{H}]\text{polyols}$ derived from $[^3\text{H}]\text{PtdIns}$ samples, the refractometer was taken out of line and the column eluant redirected to a fraction collector, and peaks of $[^{14}\text{C}]$ and $[^3\text{H}]$ determined by liquid scintillation counting.



c, Structure and derivation of adonitol from Ins(1,3)P₂.

type I PtdInsP and ^{14}C -labelled polyol standards indicated that the tritium label ran as a single peak with a mobility comparable to the poorly separating erythritol and adonitol. We distinguished between these two possibilities in the experiment illustrated in Fig. 3*b*, in which the tritiated polyol was run separately with either ^{14}C -labelled erythritol or with ^{14}C -labelled adonitol as internal standards. The ^3H -labelled polyol was clearly separable from ^{14}C -labelled erythritol, but it co-chromatographed

precisely with $[^{14}\text{C}]\text{adonitol}$. As shown in Fig. 3*c*, inositol-1,3-bisphosphate (Ins(1,3)P₂) is the only inositol bisphosphate isomer that yields adonitol after periodate oxidation, reduction, and dephosphorylation. As 60% of the HPLC-purified derived from the $[^3\text{H}]\text{InsP}_2$ type I kinase was recovered as a single peak of $[^3\text{H}]\text{adonitol}$ and no other ^3H -labelled polyol was present, we conclude that the only significant product of the type I PtdIns kinase is the novel phosphoinositide PtdIns(3)P, whereas type

II PtdIns kinase yields the conventional phosphoinositide, PtdIns(4)P.

Type I PtdIns kinase activity has been observed in a number of preparations from fibroblasts: immunoprecipitations of middle T/pp60^{c-src} from middle T-transformed NIH-3T3 cells, anti-phosphotyrosine immunoprecipitations from PDGF-stimulated BALB/C-3T3 fibroblasts, wheat germ lectin-purified material from BALB/C-3T3 fibroblasts stimulated with PDGF, and whole-cell lysates of both normal and middle T-transformed NIH-3T3 fibroblasts⁷. The enzyme obtained by these various procedures was distinguished from type II PtdIns kinase by its preference for substrate presented as vesicles rather than in detergent micelles, by its K_m for ATP, and its resistance to inhibition by adenosine⁷. We routinely use middle T/pp60^{c-src} immunoprecipitates as a convenient source of type I PtdIns kinase uncontaminated with significant type II activity. To verify that the same type of PtdIns kinase activity was present in these other preparations, their ability to phosphorylate PtdIns to PtdIns(3)P was tested. Each preparation was used to phosphorylate PtdIns with [³²P]ATP. The resulting PtdInsP was purified by TLC, deacylated and resolved by anion-exchange HPLC. The chromatographic properties of the deacylated products indicate that middle T/pp60^{c-src} immunoprecipitates produce 99% PtdIns(3)P, 1% PtdIns(4)P; anti-phosphotyrosine immunoprecipitates produce 92% PtdIns(3)P, 8% PtdIns(4)P; wheat germ lectin-precipitates produce 88% PtdIns(3)P, 12% PtdIns(4)P; cell lysates, when assayed under conditions which optimize type I PtdIns kinase activity⁷, produce 30%–50% PtdIns(3)P. The production of PtdIns(4)P by these various type I PtdIns kinase preparations presumably results from contaminating type II PtdIns kinase, but the possibility that the specificity of type I PtdIns kinase differs between these various preparations has not been excluded.

To see whether PtdIns(3)P is present in intact cells, we labelled middle T-transformed fibroblasts for 48 hours with [³H]inositol, and the phospholipids were extracted, deacylated and analysed by anion-exchange HPLC. About 95% of the deacylated PtdInsP co-eluted with ³²P-labelled GroPIns(4)P standard, whereas 3.3% of the deacylated PtdInsP co-eluted with the GroPIns(3)P standard (Fig. 4). We obtained identical results when cellular [³H]PtdInsP was purified by TLC before deacylation (data not shown). Although this suggests that PtdIns(3)P is present in intact fibroblasts, the identification of this compound requires a more rigorous analysis than is presented here. A PtdIns(3)P making up ~10% of the cellular PtdInsP pool has, however, been rigorously identified in an astrocytoma cell line (L. Stephens, P. T. Hawkins and C.P.D., manuscript in preparation). It is therefore likely that PtdIns(3)P is present in fibroblasts and is synthesized by the distinct subset of total cellular PtdIns kinase activity that we define as type I PtdIns kinase.

Our results indicate that type I PtdIns kinase can also catalyse in intact cells the reaction observed *in vitro*. Nothing is known, however, about how PtdIns(3)P is metabolized. Both fibroblast and human erythrocyte PtdInsP kinases can further phosphorylate PtdIns(3)P to PtdIns(3,x)P₂, the deacylation product of which is separable from GroPIns(4,5)P₂ by HPLC (data not shown). A possible structure for PtdIns(3,x)P₂ would be PtdIns(3,4)P₂ which would yield Ins(1,3,4)P₃ on hydrolysis by phosphoinositidase C. Ins(1,3,4)P₃ does accumulate in stimulated cells¹², but this can be accounted for by the sequential actions of Ins(1,4,5)P₃-3-kinase and Ins(1,3,4,5)P₄-5-phosphatase^{17,18}. Ins(1,3,4)P₃ could also possibly be produced directly by action of phosphoinositidase C.

PtdIns(3)P itself, rather than its metabolites, could be a critical mediator of mitogenic signals. The ability of PtdIns(3) kinase to produce an isomer of PtdInsP that is not in the pathway for Ins(1,4,5)P₃ production may be analogous to the phosphofructokinase-catalysed production of fructose-2,6-bisphosphate, which is not an intermediate in glycolysis but is the most potent known regulator of glycolytic flux¹⁷. Like the two forms of

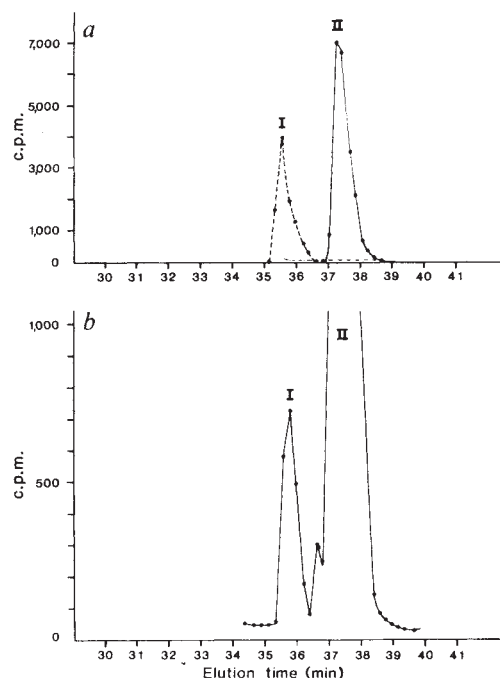


Fig. 4 PtdInsP from intact cells. Polyoma middle T-transformed NIH 3T3 fibroblasts were labelled for 72 h with 20 μ Ci ml⁻¹, [³H]inositol. Labelled phosphoinositides were extracted⁴, mixed with ³²P-labelled type I PtdInsP, deacylated and analysed by anion-exchange HPLC as described in Fig. 2 legend. *a*, Elution profile of [³²P]GroPIns(3)P co-injected with ³H-labelled deacylation products (dotted line), (I); solid line is the elution profile of [³²P]GroPIns(4)P from a separate run with ³H-labelled samples (II). *b*, elution profile of tritium counts: total ³H in peak I, 2,063 c.p.m. and in peak II, 60,500 c.p.m.

phosphofructokinase, PtdIns kinase types I and II are not isozymes, but catalyse distinct phosphorylations of a common substrate. The possibility that PtdIns(3)P is a regulatory molecule in PtdIns turnover remains to be tested experimentally, as does the involvement of type I PtdIns kinase in the action of growth factor and oncogene products. The specific association between this kinase and immunoprecipitates of polyoma middle T/pp60^{c-src} competent for transformation or of ligand-activated PDGF receptor indicates that it could be important in the transduction of signals by these tyrosine kinases.

We thank George Patten for extensive advice and assistance in the separation of intact PtdIns phosphates.

Received 23 November 1987; Accepted 18 February 1988.

- Whitman, M., Fleischman, L., Chahwala, S. B., Cantley, L. & Rosoff, P. in *PI Turnover and Receptor Function* (ed. Putney, J.) 197–217 (Liss, New York, 1986).
- Berridge, M. *Biochim. biophys. Acta* **907**, 33–45 (1987).
- Downes, C. P. & Michell, R. H. in *Molecular Aspects of Cellular Regulation Vol. 4* (eds Cohen, P. & Houslay, M. D.) 3–56 (Elsevier, Amsterdam, 1985).
- Macara, I. G. *Am. J. Physiol.* **248**, C3–C11 (1985).
- Whitman, M., Kaplan, D. R., Schaffhausen, B. S., Cantley, L. & Roberts, T. *Nature* **315**, 239–242 (1985).
- Kaplan, D. R. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 3,624–3,628 (1986).
- Whitman, M., Kaplan, D. R., Roberts, T. M. & Cantley, L. C. *Biochem. J.* **247**, 165–174 (1987).
- Kaplan, D. R. *et al. Cell* **50**, 1021–1029 (1987).
- Chang, M. & Ballou, C. E. *Biochem. biophys. Res. Commun.* **26**, 199–205 (1967).
- Brown, D. M. & Stewart, J. C. *Biochem. biophys. Acta* **125**, 413–421 (1966).
- Grado, C. & Ballou, C. E. *J. biol. Chem.* **236**, 54–60 (1961).
- Irvine, R. F., Letcher, A. J., Lander, D. J. & Downes, C. P. *Biochem. J.* **223**, 237–243 (1984).
- Prottey, C., Salway, J. G. & Hawthorne, J. N. *Biochim. biophys. Acta* **164**, 238–251 (1968).
- Endemann, G., Dunn, S. & Cantley, L. *Biochemistry* **26**, 6845–6851 (1987).
- Hawkins, P. T., Stephens, L. & Downes, C. P. *Biochem. J.* **238**, 507–516 (1986).
- Harden, T. K., Stephens, L., Hawkins, P. T. & Downes, C. P. *J. biol. Chem.* **262**, 9059–9061 (1987).
- Hers, L. & Hue, S. A. *Rev. Biochem.* **52**, 617–654 (1983).
- Hawkins, P. T., Stephens, L. & Downes, C. P. *Biochem. J.* **238**, 507–516 (1986).
- Stephens, L. R. *et al. Biochem. J.* (in the press).

GM-CSF induces human neutrophil IgA-mediated phagocytosis by an IgA Fc receptor activation mechanism

R. H. Weisbart*, A. Kacena*, A. Schuh*
& D. W. Golde†

* Department of Medicine, Veterans Administration Medical Center, Sepulveda, California 91343, USA

† UCLA School of Medicine, Los Angeles, California 90024, USA

Immunoglobulin A is the primary immunoglobulin isotype in tears, saliva, breast milk and other mucosal secretions, constituting between 6% and 15% of the total serum immunoglobulins¹. Human peripheral blood neutrophils have IgA receptors, but these cells do not normally participate in IgA-mediated phagocytosis. The haematopoietic factors granulocyte-macrophage colony-stimulating factor (GM-CSF) and granulocyte colony-stimulating factor (G-CSF) prime neutrophils to be more responsive to a variety of stimuli²⁻⁴. We therefore studied their effect on IgA-mediated phagocytosis. GM-CSF and G-CSF both induce a change from low to high-affinity neutrophil IgA Fc crystallizable fragment receptors within 30 min; a change which is associated with the development of IgA-mediated phagocytosis. Human IL-3, which does not affect neutrophil function, is inactive in this system. These results define a new mechanism for CSF-augmented host defence whereby neutrophil function can be modulated by CSF-mediated IgA Fc receptor activation.

The major function of neutrophils in host defence is phagocytosis. Neutrophil phagocytosis is enhanced by coating the particles to be ingested by immunoglobulins (opsonization), thereby providing a ligand which specifically binds to both antigen and phagocytic cells that bear immunoglobulin Fc receptors. Immunoglobulin G antibodies are efficient opsonins which enhance neutrophil-mediated phagocytosis and killing⁵. Human peripheral blood neutrophils also have surface membrane Fc receptors for IgA, but *in vitro* studies have failed to support a

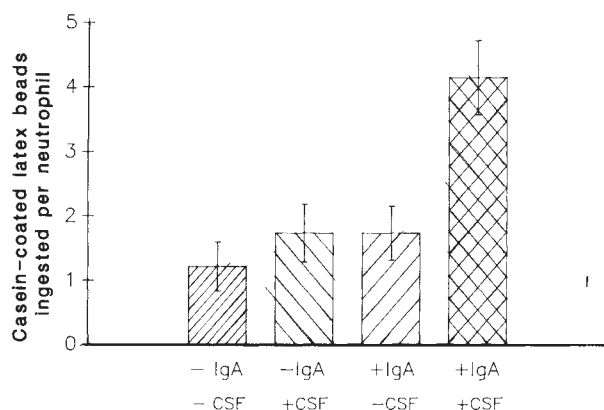


Fig. 1 Effect of GM-CSF on IgA-mediated neutrophil phagocytosis. Neutrophils were purified to >95% from human peripheral blood by Ficoll-Hypaque gradient centrifugation, and 0.1 ml aliquots containing $5 \times 10^6 \text{ ml}^{-1}$ were incubated in Hanks' balanced salt solution with 0.5% bovine serum albumin for 15 min at 37 °C in the presence and absence of 100 pM purified biosynthetic GM-CSF. Latex beads coated with casein and nitroblue tetrazolium (NBT) were incubated in the presence and absence of purified IgA (0.1 mg ml^{-1}) with immune reactivity to casein and then washed. Latex beads were added to an aliquot of neutrophils (20 beads per neutrophil) for 15 min, and each of the cell suspensions was placed in an ice bath. Aliquots were removed and observed microscopically for intracellular latex beads. The data represent the mean result of five separate experiments $\pm$ s.d.

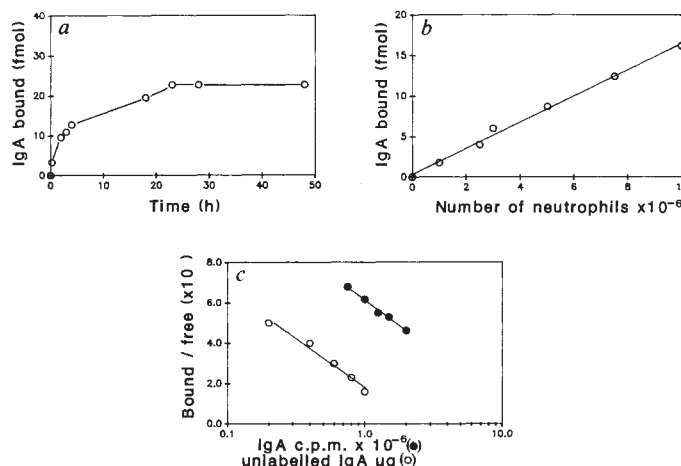


Fig. 2 Characterization of neutrophil ^{125}I -labelled IgA binding. Neutrophils were suspended in Hanks' balanced salt solution with 0.5% bovine serum albumin and incubated with ^{125}I -labelled IgA. Specific binding (IgA bound) was measured as the difference in binding in the presence and absence of excess ($5 \mu\text{g}$) unlabelled IgA. Specific binding was measured after incubation of 5×10^6 neutrophils with $0.15 \mu\text{g}$ ^{125}I -labelled IgA at 4 °C for various times to determine the time required for equilibrium binding (a). ^{125}I -labelled IgA was incubated with different numbers of neutrophils for 24 h at 4 °C to determine the effect of cell number on ^{125}I -labelled IgA binding (b). The binding curve for increasing amounts of ^{125}I -labelled IgA added to 5×10^6 neutrophils was determined (●) and compared to the binding curve obtained on adding increasing amounts of unlabelled IgA to a constant amount ($0.15 \mu\text{g}$) of ^{125}I -labelled (○) IgA in a self-displacement analysis. The slopes of the curves indicate the binding affinity of IgA and ^{125}I -labelled IgA and establish that the affinity of binding of IgA was not altered by iodination.

role for IgA-mediated neutrophil phagocytosis⁶⁻⁸. In contrast, neutrophils at mucosal sites such as those obtained from saliva can support IgA-mediated phagocytosis⁹. Thus, ligand binding to the IgA Fc receptor can stimulate a phagocytic signal under some circumstances. GM-CSF is a haematopoietic growth factor which primes mature neutrophils for enhanced oxidative metabolism in response to the major neutrophil chemoattractants^{2,3}. We previously demonstrated GM-CSF modulation of the number and affinity of f-Met-Leu-Phe receptors¹⁰. Here, we examine GM-CSF for its role in activating neutrophil IgA-mediated phagocytosis.

Human neutrophils were incubated at 37 °C for 15 min in the presence and absence 100 pM endotoxin-free biosynthetic human GM-CSF (a concentration previously shown to maximally enhance various neutrophil functions). Latex beads coated with casein and nitroblue tetrazolium (NBT) were washed and incubated in the presence and absence of purified IgA with immune reactivity to casein. The beads were added to neutrophils at 23 °C for 15 min at a ratio of 20 beads per neutrophil. The neutrophils were then examined microscopically to determine the number of ingested latex beads per 100 cells. Identification of intracellular as opposed to cell-surface associated beads was facilitated by the presence of reduced NBT. As shown in Fig. 1, neither IgA alone nor GM-CSF alone enhanced phagocytosis of the casein-coated latex beads. In contrast, neutrophils incubated with GM-CSF and then exposed to casein-coated latex beads in the presence of IgA anti-casein antibodies, showed a two to threefold increase in phagocytosis ($P=0.005$). Following treatment with GM-CSF there was an increase in the number of beads per cell. The increase in phagocytosis was inhibited (>95%) by adding an excess ($10 \mu\text{g ml}^{-1}$) of IgA without anti-casein activity to the mixture of neutrophils and casein-coated latex beads. Similar results were

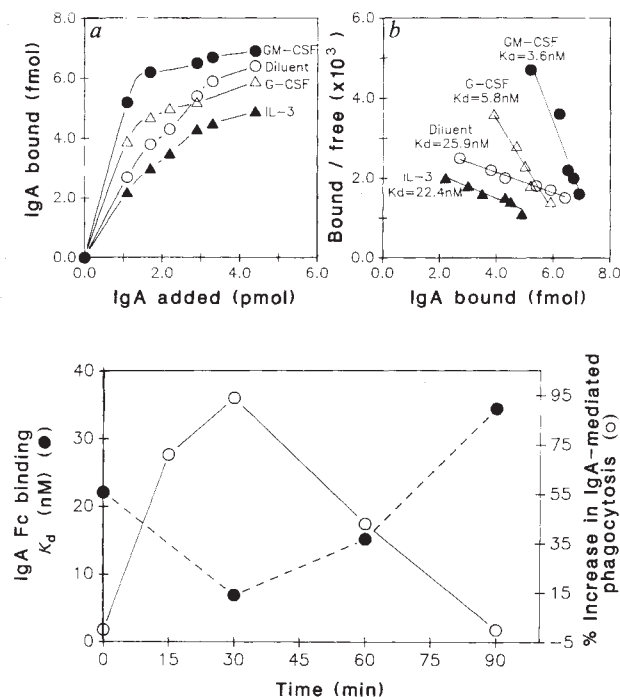


Fig. 3 Effects of GM-CSF and G-CSF on neutrophil IgA receptor binding. Neutrophils were incubated for 30 min at 37 °C with diluent plus 100 pM GM-CSF, G-CSF or IL-3. Increasing amounts of ¹²⁵I-labelled IgA were added to aliquots of 5 × 10⁶ neutrophils for 24 h at 4 °C in the presence and absence of excess unlabelled IgA under conditions of equilibrium binding. Bound ¹²⁵I-labelled IgA was separated from free IgA by centrifuging the neutrophils through a gradient of paraffin oil and silicon oil¹⁰. The results are expressed as IgA neutrophil binding (a) and by Scatchard analysis (b). Binding affinities are indicated by the dissociation constants (K_d).

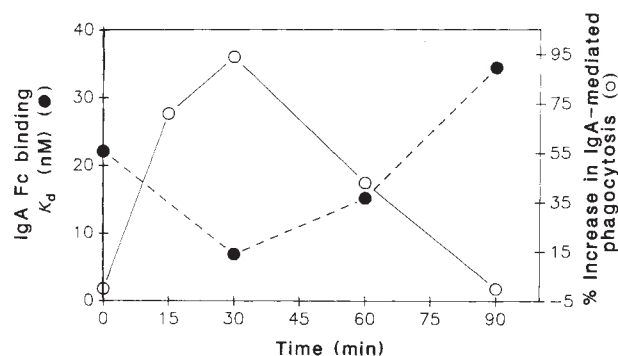


Fig. 4 Neutrophil responses to GM-CSF. The time course for the effect of GM-CSF on both IgA-mediated phagocytosis and IgA receptor binding affinity is shown. Phagocytosis of casein-coated latex beads (○) was measured as described in Fig. 1. The results are expressed as % increase in IgA-mediated phagocytosis. Binding of ¹²⁵I-labelled IgA (●) was determined by Scatchard analysis as described in Fig. 3.

obtained in response to 100 pM biosynthetic human G-CSF with a threefold increase in phagocytosis. These findings demonstrate that GM-CSF and G-CSF activate human neutrophil IgA-mediated phagocytosis.

To ascertain possible changes in the IgA receptor associated with GM-CSF-activated IgA-mediated neutrophil phagocytosis, the number and affinity of neutrophil IgA receptors was determined in the presence and absence of GM-CSF. Purified IgA was iodinated by the method of Bolton and Hunter and the conditions for equilibrium binding to human peripheral blood neutrophils were established (Fig. 2). The binding of ¹²⁵I-labelled IgA was also shown to be linearly related to cell number (Fig. 2). The specific activity of ¹²⁵I-labelled IgA preparations as determined by self-displacement analysis (Fig. 2) varied from 3.3 × 10⁶ c.p.m. per μg IgA to 6.7 × 10⁶ c.p.m. per μg IgA, and 50–60% of ¹²⁵I-labelled IgA binding was inhibited by unlabelled IgA^{11,12}. The self-displacement analysis also showed that iodination did not change the binding affinity of the IgA (Fig. 2).

Saturation binding of ¹²⁵I-labelled IgA to neutrophils was demonstrated (Fig. 3a), and the Scatchard analysis of one representative experiment (Fig. 3b) demonstrates an increase in affinity but a decrease in number of IgA receptors in response to GM-CSF and G-CSF. The limited number of data points does not permit evaluation of the coexistence and relative expression of low and high affinity receptors. In eight separate Scatchard plot experiments, the mean K_d ± s.d. for IgA binding to control (diluent-treated) neutrophils was 21.0 ± 7.2 nM. In contrast, neutrophils incubated for 30 min with physiological concentrations of GM-CSF or G-CSF (100 pM) showed a marked increase in affinity. In six separate Scatchard plot experiments the mean K_d ± s.d. for IgA in response to 100 pM GM-CSF was 6.3 ± 3.6 nM. In contrast to GM-CSF and G-CSF, human IL-3 (100 pM) did not significantly alter the affinity of neutrophil IgA Fc receptors.

The time course of action of GM-CSF on both neutrophil phagocytosis and the increase in affinity of neutrophil IgA binding was similar, with maximal responses occurring between 15 and 30 min (Fig. 4). Responses were no longer evident after GM-CSF was incubated with neutrophils for 90 min, and the number of low affinity receptors returned to pretreatment levels.

The significance of neutrophil Fc receptors for IgA in the absence of IgA-mediated phagocytosis has been an enigma,

especially as peripheral blood monocytes and a subpopulation of T lymphocytes have IgA Fc receptors which can mediate IgA-dependent functional activities^{13,14}. The development of IgA-mediated phagocytosis by mucosal neutrophils, however, suggested that neutrophil IgA receptors can be modified to become functionally active⁹. Moreover, cells of the human promyelocytic line HL-60 were recently shown to develop IgA receptors and IgA-mediated phagocytosis after treatment with calcitriol, a differentiation-inducing agent¹⁵. Haematopoietic factors are logical candidates for activators of neutrophil IgA receptors. Our studies show that both GM-CSF and G-CSF but not IL-3 induce the expression of high affinity neutrophil IgA receptors. Concomitant with the development of high affinity IgA Fc receptors, neutrophils display enhanced phagocytosis mediated by IgA.

These results identify a new physiological role for GM-CSF and G-CSF in regulating neutrophil host defence function and also define a novel mechanism for receptor activation, thus permitting IgA-mediated phagocytosis by neutrophils.

We thank Steven C. Clark (Genetics Institute Inc., Cambridge, Massachusetts) for providing IL-3, Fumimaru Takaku for providing human G-CSF, and Judith Gasson (UCLA) for helpful discussions. This work was supported by USPHS grants, a grant from the Veterans Administration, and a grant from the Southern California Chapter of the Arthritis Foundation.

Received 8 January 1988; accepted 19 February 1988.

- Conley, M. E. & Delacroix, D. L. *Ann. intern. Med.* **106**, 892–899 (1987).
- Weisbart, R. H., Golde, D. W., Clark, S. C., Wong, G. C. & Gasson, J. C. *Nature* **314**, 361–363 (1985).
- Weisbart, R. H., Kwan, L., Golde, D. W. & Gasson, J. C. *Blood* **69**, 18–21 (1987).
- Lopez, A. F. et al. *J. Immunol.* **131**, 2983–2988 (1981).
- Klebanoff, D. J. & Clark, R. A. *The Neutrophil: Function and Clinical Disorders* (North-Holland, Amsterdam, New York, Oxford, 1978).
- Quie, P. G., Messner, R. P. & Williams, R. C. Jr *J. exp. Med.* **128**, 553–570 (1968).
- Wilson, J. D. *J. Immunol.* **108**, 726–730 (1972).
- Zipursky, A., Brown, E. J. & Bienenstock, J. *Proc. Soc. exp. Biol. (N.Y.)* **142**, 181–184 (1973).
- Fanger, M. W., Goldstine, S. N. & Shen, L. *Molec. Immunol.* **20**, 1019–1027 (1983).
- Weisbart, R. H., Golde, D. W. & Gasson, J. C. *J. Immunol.* **137**, 3584–3587 (1986).
- Gasson, J. C., Kaufman, S. E., Weisbart, R. H., Tomonaga, M. & Golde, D. W. *Proc. natn. Acad. Sci. U.S.A.* **83**, 669–673 (1986).
- Calvo, J. C., Radice, J. P. & Charreau, E. H. *Biochem. J.* **212**, 259–264 (1983).
- Tagliabue, A. et al. *J. Immunol.* **133**, 988–992 (1984).
- Tagliabue, A., Villa, L., Boraschi, K., Peri, G., deGori, V. & Nencioni, L. *J. Immunol.* **135**, 4178–4182 (1985).
- Maliszewski, C. R., Shen, L. & Fanger, M. W. *J. Immunol.* **135**, 3878–3881 (1985).

The JUN oncoprotein, a vertebrate transcription factor, activates transcription in yeast

Kevin Struhl

Department of Biological Chemistry, Harvard Medical School, Boston, Massachusetts, 02115, USA

Transcriptional activation of RNA polymerase II in eukaryotic organisms ranging from yeasts to mammals has many common features such as enhancer elements, TATA elements, and activator proteins that bind specifically to promoter DNA (reviewed in refs (1, 2)). The JUN oncoprotein, which causes sarcomas in chickens³, shows significant homology to the DNA-binding domain of GCN4, a yeast protein that stimulates transcription of the amino acid biosynthetic genes⁴. The GCN4 and JUN proteins bind the same DNA sequences⁵, consensus ATGA(C/G)TCAT (ref. 6), even though the DNA-binding domains are only 45% identical in amino acid sequence. The JUN protein almost certainly represents the oncogenic version of the normal AP-1 transcription factor⁷, suggesting an evolutionary relationship between yeast and vertebrate activator proteins. Here, I demonstrate that JUN efficiently activates transcription in yeast either through its own or a heterologous DNA-binding domain. As is the case for yeast activator proteins, transcriptional stimulation by JUN requires an acidic activation region distinct from the DNA-binding domain. The functional interchangeability between yeast and vertebrate transcription factors strongly suggests a basic similarity in the molecular mechanism of eukaryotic transcriptional activation.

GCN4 binds specifically to the promoters of many genes involved in amino acid biosynthesis and coordinately induces their transcription^{8,9}. GCN4 contains 281 amino acids and binds as a dimer¹⁰ with optimal binding to the 9 base pair dyad ATGA(C/G)TCAT (ref. 6). The 60 C-terminal amino acids of GCN4 are sufficient both for specific DNA-binding¹¹ and for dimerization¹⁰. Although the GCN4 DNA-binding domain is necessary for recognizing the appropriate promoters, a short acidic region in the centre of the region is required for transcriptional activation¹¹. This acidic activation region is also capable of transcriptional stimulation when fused to a heterologous DNA-binding domain, the *E. coli* LexA repressor. The resulting LexA-GCN4 hybrid protein is a bifunctional activator because it activates transcription from promoters containing either GCN4 or LexA binding sites upstream of TATA sequences¹¹.

Previously, I described a LexA-GCN4-JUN hybrid protein (originally called LGJ-1) in which the entire GCN4 DNA-binding domain (the C-terminal 112 amino acids) was replaced by the 166 C-terminal amino acids of JUN⁵. When introduced into yeast cells lacking the entire *GCN4* gene, this protein induced the transcription of *HIS3* and other amino acid biosynthetic genes, thus indicating that the JUN and GCN4 DNA-binding domains are functionally homologous. However, because the LexA-GCN4-JUN hybrid protein contained the intact GCN4 activation region, this experiment did not bear on the issue of whether the JUN protein itself was capable of transcriptional activation in yeast.

To address this question, I constructed a LexA-JUN hybrid protein in which essentially the entire JUN coding region was fused directly to the LexA DNA-binding domain (Fig. 1). A plasmid capable of expressing this protein was introduced into KY330, a yeast strain lacking the entire *GCN4* gene and harbouring a separate plasmid that contains the *E. coli* LexA repressor fused upstream of a yeast TATA element and β -galactosidase structural gene. In this way, transcriptional activation through the LexA DNA-binding domain could be quantified by measuring β -galactosidase activity, and activation through the

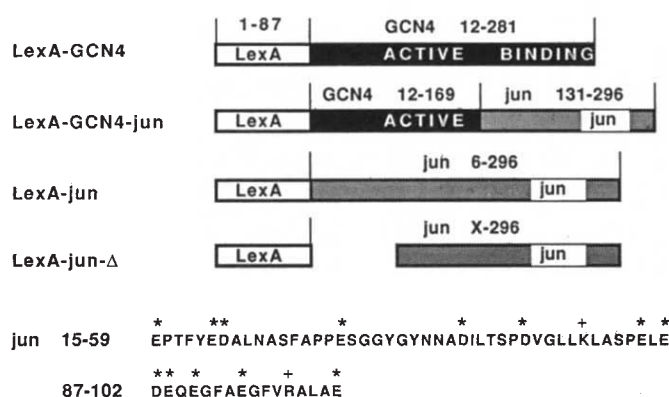


Fig. 1 Structures of LexA hybrid proteins, composed of the LexA DNA-binding domain (open box), the GCN4 coding region (black box with transcriptional activation and DNA-binding domains indicated), and the JUN coding region (shaded box with the DNA-binding domain indicated by JUN). The constructions of LexA-GCN4¹¹ and LexA-GCN4-JUN⁵ have been described previously. To generate the DNA molecule encoding LexA-JUN, *jun* plasmid DNA was partially cleaved with *Fnu*DI, ligated to an octanucleotide *Sal*I linker, cleaved with *Sal*I and *Eco*RI, and the appropriate fragment fused to DNA encoding the LexA domain. Due to the cloning procedure, there are five additional amino acids located at the junction between the LexA and JUN coding regions. LexA-JUN deletion mutants were generated similarly except that JUN plasmid DNA was treated with *Bal*31 nuclease prior to ligating the *Sal*I linker. The end-points within the JUN coding sequence were determined by DNA sequencing. All protein coding regions are cloned in YCp88, a vector containing the *URA3* marker, *ARS1* and *CEN3* elements for maintenance as a minichromosome at one copy per cell, and the *DED1* promoter for expression of the hybrid protein¹¹. The acidic regions of JUN are shown, acidic residues being marked * and basic residues +.

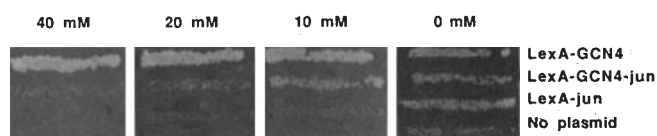


Fig. 2 Phenotypic analysis. YCp88 plasmids encoding the LexA hybrid proteins (Fig. 1) or no protein were introduced into yeast strain KY330 by selecting for *Ura*⁺ colonies. Total GCN4 function was assayed by complementation of the *gcn4*-Δ1 allele by plating the resulting transformants in minimal media lacking leucine and containing the indicated concentrations of aminotriazole, a competitive inhibitor of the *HIS3* gene product. The degree of aminotriazole resistance is directly related to the level of transcription of *HIS3* and other amino acid biosynthetic genes⁶, normally regulated by GCN4.

JUN DNA-binding domain could be assessed by complementation of the *gcn4* deletion. The complementation assay was based on the facts that *gcn4* deletion strains grow slowly on minimal medium lacking amino acids and fail to grow in the presence of aminotriazole, a competitive inhibitor of the *HIS3* gene product. These phenotypes reflect the inability of *gcn4* deletion strains to induce transcription of *HIS3* and other amino acid biosynthetic genes above the basal level.

The basic result is that LexA-jun activates transcription thus indicating that the JUN oncogene contains a sequence(s) that functions as a transcriptional activation region in yeast. LexA-JUN can functionally replace GCN4 even though this hybrid protein contains none of the GCN4 coding region. KY330 cells expressing LexA-JUN grew in the presence of 10 or 20 mM aminotriazole, and grew essentially at wild-type rates in minimal medium lacking amino acids (Fig. 2). In addition, these cells

also show a high level of β -galactosidase activity (Table 1). Thus, LexA-JUN activates transcription through either its own DNA-binding domain or the heterologous LexA DNA-binding domain.

The level of transcriptional activation by LexA-JUN was compared with that achieved by LexA-GCN4-JUN, which contains the GCN4 transcriptional activation region and the JUN DNA-binding domain, and LexA-GCN4, which contains essentially the entire GCN4 coding region (Fig. 1). When examined for activation through the LexA domain, all three proteins confer comparable levels of β -galactosidase (Table 1). This suggests that JUN and GCN4 contain transcriptional activation regions that are equally functional when fused to a heterologous DNA-binding domain.

When assayed for activation through the JUN DNA-binding domain, LexA-JUN appears to be slightly less effective than LexA-GCN4-JUN in that cells grow more slowly in the presence of aminotriazole (Fig. 2). This effect is rather subtle: using the well-established correlation between the degree of aminotriazole resistance and the level of *HIS3* messenger RNA⁶, it can be estimated that the level of activation by these proteins differs by a factor of <2. Both proteins containing the JUN DNA-binding domain complement the *gcn4* deletion less efficiently than LexA-GCN4; that is, they activate transcription of the amino acid biosynthetic genes to lower levels. As discussed previously⁵, this probably reflects the relative affinities of the GCN4 and JUN DNA-binding domains to their target sites *in vivo*. Nevertheless, in terms of the transcriptional activation function, the JUN oncoprotein is nearly as functional in yeast cells as GCN4, a native yeast activator protein.

Deletion analyses of the yeast GCN4 and GAL4 activator proteins have shown that transcriptional activation regions are defined by short acidic regions with no primary sequence homology^{1,11,12}. Indeed, JUN contains a 45 amino acid region between residues 15 and 59 with a net negative charge of -7 and a 16 amino acid region between residues 87 and 102 with a net charge of -4 (ref. 3, fig. 1). As shown in Table 1, deletions that remove more than 100 N-terminal residues of JUN and hence lack both acidic regions confer extremely low levels of activation. Deletions with end-points between residues 54 and 71, which remove one of the acidic regions show a 2-5 fold decrease in the level of expression. Thus, this initial deletion analysis of LexA-JUN indicates that these acidic regions are important for transcriptional activation in yeast.

The observation that approximately 1% of short *E. coli* sequences act as a transcriptional activation region when fused to a 147 amino acid GAL4 DNA-binding domain¹³, raises the formal possibility that JUN activation in yeast cells may not reflect a meaningful functional relationship between yeast and vertebrate transcription factors but rather the presence of a sequence that fortuitously behaves as a yeast activation region. However, this possibility is very unlikely because JUN activates transcription through its own or the LexA DNA-binding domain at levels that are near those achieved by the intact GCN4 activation region. In contrast, essentially all the 'functional' *E. coli* segments activate transcription poorly through the GAL4 DNA-binding domain and not at all through the LexA domain¹³. In addition, after fusing random DNA sequences to the LexA DNA-binding domain, we were unable to select for transcriptional activators out of a population of approximately 10⁴ LexA hybrid proteins (A. R. Oliphant and K.S., unpublished observations). Thus it is rare to obtain protein sequences that can stimulate transcription when fused to the LexA domain, especially at levels that are comparable to that obtained with a native yeast activation region. It is also worth noting that the 147-residue GAL4 DNA-binding domain used for identifying the *E. coli* activation segments contains about 70 additional residues beyond those necessary for DNA binding including an acidic region (net negative charge of -7 between residues 65 to 117). Thus the *E. coli* peptides might not be independent activa-

Table 1 Activation through the LexA DNA-binding domain.

Protein	Activation region	<i>jun</i> sequences	β -galactosidase
LexA-GCN4	GCN4	None	410
LexA-GCN4-JUN	GCN4	130-296	310
LexA-JUN	JUN	6-296	380
LexA-JUN- Δ 1	JUN	22-296	300
LexA-JUN- Δ 2	JUN	54-296	170
LexA-JUN- Δ 3	JUN	68-296	110
LexA-JUN- Δ 4	JUN	71-296	140
LexA-JUN- Δ 5	JUN	101-296	10
LexA-JUN- Δ 6	JUN	124-296	8
LexA-JUN- Δ 7	JUN	136-296	5
LexA-JUN- Δ 8	JUN	137-296	5
LexA-JUN- Δ 9	JUN	147-296	2
LexA-gcn4-N77	None	None	<1
None	None	None	<1

Yeast strain KY330, a derivative of KY803 (*trp1- Δ 1 ura3-52 leu2-P1 gcn4- Δ 1*)¹¹ harbouring the YEp21-Sc3423 plasmid containing a *lexA* operator and *cyc1* TATA element upstream of the *lacZ* structural gene¹¹, was transformed by YCp88 plasmids capable of expressing various LexA hybrid proteins (see Figs. 1, 2). The resulting transformants were grown in minimal medium lacking leucine and assayed for β -galactosidase activity as described previously¹¹. The listed values represent the averages of three independent determinations and have an error of approximately $\pm 20\%$.

tion regions, but rather inactive or partially active segments that act in combination with the otherwise cryptic GAL4 acidic region. These considerations strongly support the idea that JUN activation in yeast reflects both an evolutionary and mechanistic similarity between yeast and vertebrate transcription factors.

It has been suggested that yeast activator proteins such as GCN4 and GAL4 stimulate transcription by contacting components of the basic transcription machinery¹¹⁻¹⁵, possibly proteins that bind to the highly conserved TATA elements¹. The results presented here suggest that JUN, an oncogenic version of the normal AP-1 transcriptional factor of vertebrates⁷, can interact functionally with the basic transcription machinery of yeast. The obvious implication is that the basic transcription machineries of eukaryotic organisms from yeast to man are evolutionarily conserved. In support of this idea, it has been shown very recently that GAL4 activates transcription in mammalian cells^{16,17}, and that the FOS oncoprotein, a hypothetical transcription factor, activates transcription in yeast cells¹⁸. Such functional interchangeability may make it fruitful to study the mechanism of transcriptional activation *in vivo* or *in vitro* using mixtures of yeast and mammalian components. Thus, contrary to some beliefs, it appears that mRNA transcriptional initiation in all eukaryotes may occur by a common molecular mechanism.

I thank Pehr A. B. Harbury for excellent technical assistance. This work was supported by a grant from the National Institutes of Health.

Received 19 January; accepted 3 March 1988.

- Struhl, K. *Cell* **49**, 295-297 (1987).
- McKnight, S. L. & Tjian, R. *Cell* **46**, 795-805 (1986).
- Maki, Y., Bos, T. J., Davis, C., Starbuck, M., & Vogt, P. K. *Proc. natn. Acad. Sci. U.S.A.* **84**, 2848-2852 (1987).
- Vogt, P. K., Bos, T. J., & Doolittle, R. F. *Proc. natn. Acad. Sci. U.S.A.* **84**, 3316-3319 (1987).
- Struhl, K. *Cell* **50**, 841-846 (1987).
- Hill, D. E., Hope, I. A., Macke, J. P. & Struhl, K. *Science* **234**, 451-457 (1986).
- Bohmann, D., Bos, T. J., Admon, A., Nishimura, T., Vogt, P. K. & Tjian, R. *Science* **238**, 1386-1392 (1987).
- Hope, I. A. & Struhl, K. *Cell* **43**, 177-188 (1985).
- Arndt, K. & Fink, G. *Proc. natn. Acad. Sci. U.S.A.* **83**, 8516-8520 (1986).
- Hope, I. A. & Struhl, K. *EMBO J.* **6**, 2781-2784 (1987).
- Hope, I. A. & Struhl, K. *Cell* **46**, 885-894 (1986).
- Ma, J. & Ptashne, M. *Cell* **48**, 847-853 (1987).
- Ma, J. & Ptashne, M. *Cell* **51**, 113-139 (1987).
- Brent, R. & Ptashne, M. *Cell* **43**, 729-736 (1985).
- Chen, W., Tabor, S. & Struhl, K. *Cell* **50**, 1047-1055 (1987).
- Kakidani, H. & Ptashne, M. *Cell* **52**, 161-167 (1988).
- Webster, N., Jin, K. R., Green, S., Hollis, M. & Chambon, P. *Cell* **52**, 16
- Lech, K., Anderson, K. & Brent, R. *Cell* **52**, 179-184 (1988).

The origin of MHC class II gene polymorphism within the genus *Mus*

Thomas J. McConnell*, William S. Talbot,
Richard A. McIndoe & Edward K. Wakeland

Laboratory of Molecular Biology, Department of Pathology,
University of Florida, Gainesville, Florida 32610, USA

The I region of the major histocompatibility complex (MHC) of the mouse (*H-2*) contains a tightly-linked cluster of highly polymorphic genes (class II MHC genes) which control immune responsiveness. Speculation on the origin of this polymorphism, which is believed to be essential for the function of the class II proteins in immune responses to disease, has given rise to two hypotheses. The first is that hypermutational mechanisms (gene conversion or segmental exchange) promote the rapid generation of diversity in MHC genes. The alternative is that polymorphism has arisen from the steady accumulation of mutations over long evolutionary periods, and multiple specific alleles have survived speciation (trans-species evolution). We have looked for evidence of 'segmental exchange' and/or 'trans-species evolution' in the class II genes of the genus *Mus* by molecular genetic analysis of *I-A_B* alleles. The results indicate that >90% (28 out of 31) of the alleles examined can be organized into two evolutionary groups both on the basis of restriction site polymorphisms and by the presence or absence of a short interspersed nucleotide element (SINE). Using this SINE sequence as an evolutionary tag, we demonstrate that *I-A_B* alleles in these two evolutionary groups diverged at least three million years ago and have survived the speciation events leading to several modern *Mus* species. Nucleotide sequence comparisons of eight *Mus m. domesticus* *I-A_B* alleles representing all three evolutionary groups indicate that most of the divergence in exon sequences is due to the steady accumulation of mutations that are maintained independently in the different alleles. But segmental exchanges between alleles from different evolutionary groups have also played a role in the diversification of β_1 exons.

Speciation within *Mus* has been extensively studied and several closely related species have been characterized^{7-9,19}.

* Present address: Biological Resources Branch Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA.

These include the commensal species *Mus m. musculus*, *Mus m. domesticus*, and *Mus m. castaneus* which diverged ~1-2 million years ago, and the aboriginal species *Mus hortulanus* and *Mus spretus*, which diverged from commensal *Mus musculus* ~3-4 million years ago. We assembled a collection of 49 *H-2* haplotypes derived from these five *Mus* species¹⁰⁻¹² and defined 31 *I-A_B* alleles by analysis of restriction fragment length polymorphisms (RFLPs). The degree of genomic sequence divergence among these 31 alleles was estimated by calculating the fraction of restriction fragments (*F*) and sites shared (*S*) by pairs of alleles¹³. This analysis ordered the 31 *I-A_B* alleles into three distinct evolutionary groups (Table 1). Most such alleles (28 out of 31) were in either evolutionary group 1 or 2, both of which contained alleles derived from four separate *Mus* species. These findings are consistent with the trans-species evolution of *I-A_B* and contrast with data obtained when other nuclear genes or mitochondrial DNA polymorphisms were analysed in mice from these populations^{7-9,14}.

Genomic sequence comparisons of *I-A_B*^a (group 1) and *I-A_B*^b (group 2)^{15,16} show that the region of highest divergence between these alleles occurs in the intron separating the β_1 and β_2 exons (Fig. 1a). The sequence homology in this intron is <90%, and *I-A_B*^b contains 861 base pairs (bp) of inserted sequence which consists entirely of short interspersed nucleotide elements. SINEs, commonly termed retroposons, are sequences that integrate randomly throughout the genome by a poorly defined mechanism which may involve an RNA intermediate^{17,18}.

The relationship of this retroposon polymorphism to the evolutionary groups defined by RFLP analysis was tested by genomic restriction mapping of all the *I-A_B* alleles in groups 1 and 2 (Fig. 1b, c). The retroposon is characteristic of group 2 alleles. It is not the only distinguishing feature, however, since a pairwise analysis of shared restriction sites excluding the retroposon sequence demonstrates that alleles in each group share group-specific restriction sites (Table 1). These data indicate that alleles in groups 1 and 2 have clearly distinguishable intron structures.

Segmental exchange events have been postulated to play an important role in the diversification of exon sequences in class II genes². Evidence of segmental exchanges affecting these exons can be obtained by comparing the evolutionary relationships of both exons and introns between alleles. If *I-A_B* diversity only reflects the steady accumulation of mutations for millions of years, then *I-A_B* alleles should evolve as a single unit and exon

Table 1 Definition of three evolutionary groups of *I-A_B* alleles by quantitative RFLP analysis

Evolutionary group	No. of <i>I-A_B</i> alleles	<i>F</i> values		<i>S</i> values		<i>Mus</i> species	<i>H-2</i> haplotypes
		within group	between group	within group	between group		
1	14	0.65 ± 0.02	0.15 ± 0.01	0.84 ± 0.01	0.76 ± 0.01	<i>domesticus</i>	d, p, q, r, v, nod*, w3, w5, w11, w14, w22, w28, w29, w31, w217, w218
						<i>musculus</i>	w220, w222, w224, w225
						<i>castaneus</i>	w17, w228, w229
						<i>hortulanus</i>	w231
2	14	0.66 ± 0.01	0.15 ± 0.01	0.81 ± 0.01	0.76 ± 0.02	<i>domesticus</i>	b, f, j, s, w15, w233, w34, w201, w206, w207, w208, w215, w216, w219, w30, w31, t ^{w12}
						<i>musculus</i>	w221, w71
						<i>hortulanus</i>	w237
						<i>spretus</i>	w230
3	3	0.66 ± 0.02	0.13 ± 0.02	ND	ND	<i>domesticus</i>	k, u, z, w26

A total of 49 independently-derived *H-2* haplotypes were analysed, 13 from standard laboratory mice and 36 from wild mice^{7,12,24}. *F* values were calculated¹¹ from the results of RFLP analysis with restriction endonucleases *Bam*HI, *Eco*RI, *Hind*III, *Pst*I, *Pvu*II, *Bgl*II and *Sac*I. The mean ± s.e.m. values obtained in a pairwise comparison of all *I-A_B* alleles within the same group, or all *I-A_B* alleles in one group with all *I-A_B* alleles in other groups are presented. *S* values were calculated as the proportion of shared restriction sites in a pairwise analysis between alleles in groups 1 and 2, excluding all sites within the retroposon insertion. The mean *S* value ± s.e.m. from comparisons of alleles within the same group or between groups is presented. The difference between the mean *S* value within groups versus between groups was statistically significant in both comparisons (*P* < 0.001). ND, not determined.

* The RFLP analysis and restriction map of *I-A_B*^{nod} is described in ref. 25.

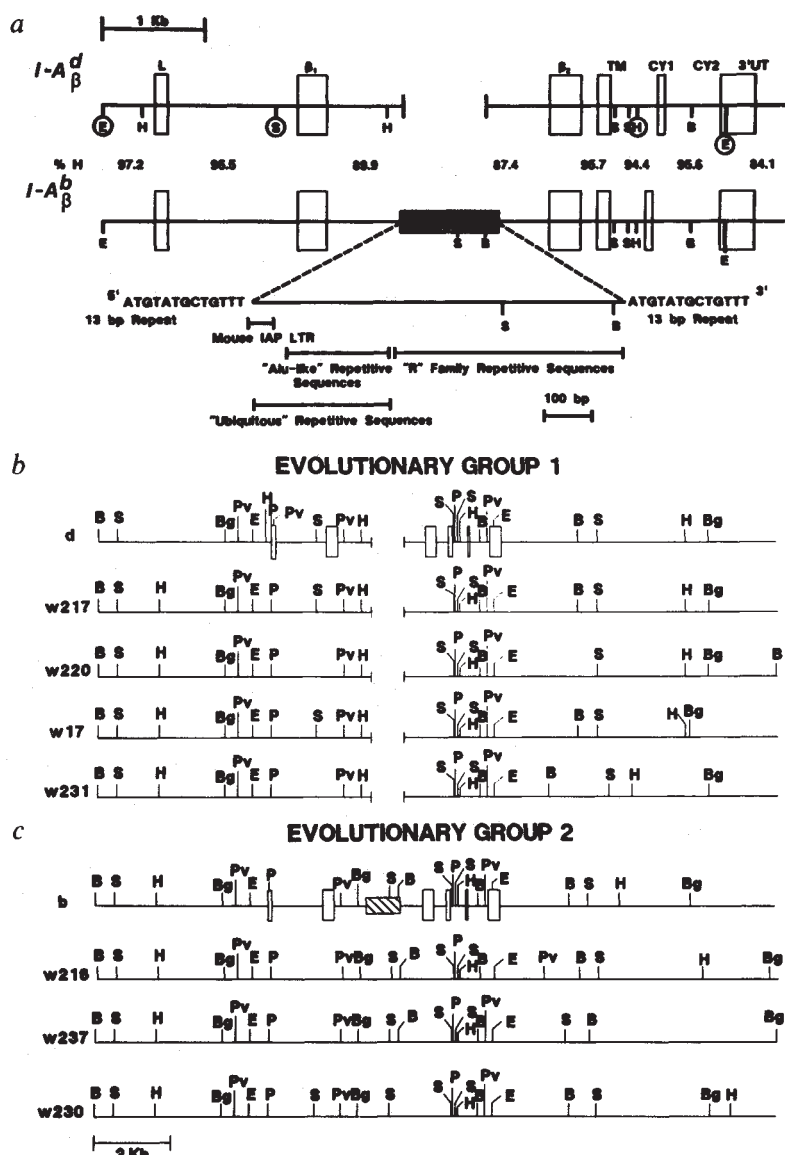


Fig. 1 Nucleotide sequence and restriction mapping analyses of *I-A_β* alleles in evolutionary groups 1 and 2. **a**, The sequence homologies (%H) of the introns and flanking regions of *I-A_β^d* (group 1) and *I-A_β^b* (group 2) were calculated from sequences available in the Genetic Sequences Data Bank (Genbank) using the MicroGenie sequence software package (Beckman). The exons of each gene are identified by open boxes and labelled on the basis of the protein domains they encode. The hatched box in *I-A_β^b* represents an 861 bp sequence which was absent in *I-A_β^d*. An expanded version of this inserted sequence is presented below, together with the 13 bp host-derived flanking direct repeats. Regions of homology with short interspersed nucleotide elements (SINES) were revealed on searching in Genbank and are indicated beneath the insert. **b**, **c**, Representative restriction maps of *I-A_β* alleles in evolutionary groups 1 and 2. Open boxes represent the exons as labelled in **a**. The gaps in the maps of group 1 alleles indicate the site of the retroposon (shaded box) in evolutionary group 2 alleles. Restriction endonuclease sites are labelled as follows: B, *Bam*HI; S, *Sac*I; H, *Hind*III; Bg, *Bgl*II; Pv, *Pvu*II; E, *Eco*RI; P, *Pst*I. The alleles presented are derived from the following *Mus* species: *I-A_β^d*, laboratory strain (probably *Mus m. domesticus*); *I-A_β^{w217}*, *Mus m. domesticus*; *I-A_β^{w220}*, *Mus m. musculus*; *I-A_β^{w17}*, *Mus m. castaneus*; *I-A_β^{w231}*, *Mus hortulanus*; *I-A_β^b*, laboratory strain (probably *Mus m. domesticus*); *I-A_β^{w216}*, *Mus m. domesticus*; *I-A_β^{w237}*, *Mus hortulanus*; *I-A_β^{w230}*, *Mus spretus*.

Methods. Restriction fragments were ordered by sequential re-hybridizations of the nylon membranes used for RFLP analyses with probes specific for either the 5' or 3' regions of *I-A_β*. Fragments derived from the 5' end of *I-A_β* were detected with the *Eco*RI-*Sac*I fragment which includes the leader sequence and 3' derived fragments were detected with the *Hind*III-*Eco*RI fragment which terminates in the 3' untranslated region (relevant restriction sites are circled in the map of *I-A_β^d*, panel **a**). The positions of polymorphic restriction sites were determined by appropriate digestions with multiple restriction enzymes (data not shown).

sequence polymorphisms should remain linked to intron sequence polymorphisms. In contrast, segmental exchange events operating specifically on exons would dissociate the evolution of those exon sequence polymorphisms from intron polymorphisms.

Evidence of segmental exchange events was sought by comparing the published exon sequences of eight *I-A_β* alleles^{15,16,20,21} derived from different evolutionary groups (Fig. 2). Sequence homologies of β_2 exons from alleles in the same group are $99.1 \pm 0.2\%$, versus $96.6 \pm 0.3\%$ between groups ($P < 0.001$, Fig. 2b), and 10 of the 19 sequence polymorphisms are group-specific. These results indicate that all eight β_2 exon sequences have evolved in association with the intron sequence polymorphisms which define *I-A_β* evolutionary groups. A similar analysis of the β_1 exons (Fig. 2a) demonstrates that only six of the eight have sequence polymorphisms which point to their evolutionary group assignments. The average sequence homology of these six β_1 exons is $95.5 \pm 0.6\%$ within groups versus $91.1 \pm 0.4\%$ between groups ($P < 0.001$), and 20 of the 56 sequence polymorphisms are group specific (boxed in Fig. 2a).

The two alleles with β_1 exon sequences that do not reflect their evolutionary group assignments, *A_β^b* and *A_β^{nod}*, appear to be the products of intragenic segmental exchange events. *I-A_β^b* has intron and β_2 exon sequences characteristic of group 2

alleles but a β_1 exon which is more closely homologous to group 1 alleles (95.1 ± 0.1 with group 1 versus $89.9 \pm 0.9\%$ with group 2, $P < 0.01$). In *I-A_β^{nod}* the 5' portion of the β_1 exon is consistent with its assignment as a group 1 allele but the 3' portion is more homologous to β_1 exons in group 3 (Fig. 2a).

These findings establish an evolutionary framework with which to interpret the polymorphism of *I-A_β*. The evolutionary divergence of the three groups defined by the retroposon polymorphisms is depicted in Fig. 3a. Group 2 *I-A_β* alleles arose from a retroposon insertion into a group 1 *I-A_β* allele at least 3–4 million years ago. Subsequently, alleles with and without the retroposon persisted through several speciations and both forms are present in a number of modern *Mus* species. These findings are consistent with the maintenance of MHC diversity over long evolutionary periods by strong selective pressures²⁶, and exclude the possibility that all the allelic diversity of *I-A_β* has been generated subsequent to speciation.

Our interpretation of the evolution of the β_1 and β_2 exons is shown schematically in Fig. 3b. The nucleotide sequences of most β_1 and all β_2 exons reflect the evolutionary origin of the allele concerned, indicating that the polymorphism observed is predominantly due to the steady accumulation of mutations in *I-A_β* during trans-species evolution. The β_1 exon sequences in the *I-A_β^b* and *I-A_β^{nod}* alleles, however, indicate that segmental exchange mechanisms also play a role in the diversification of

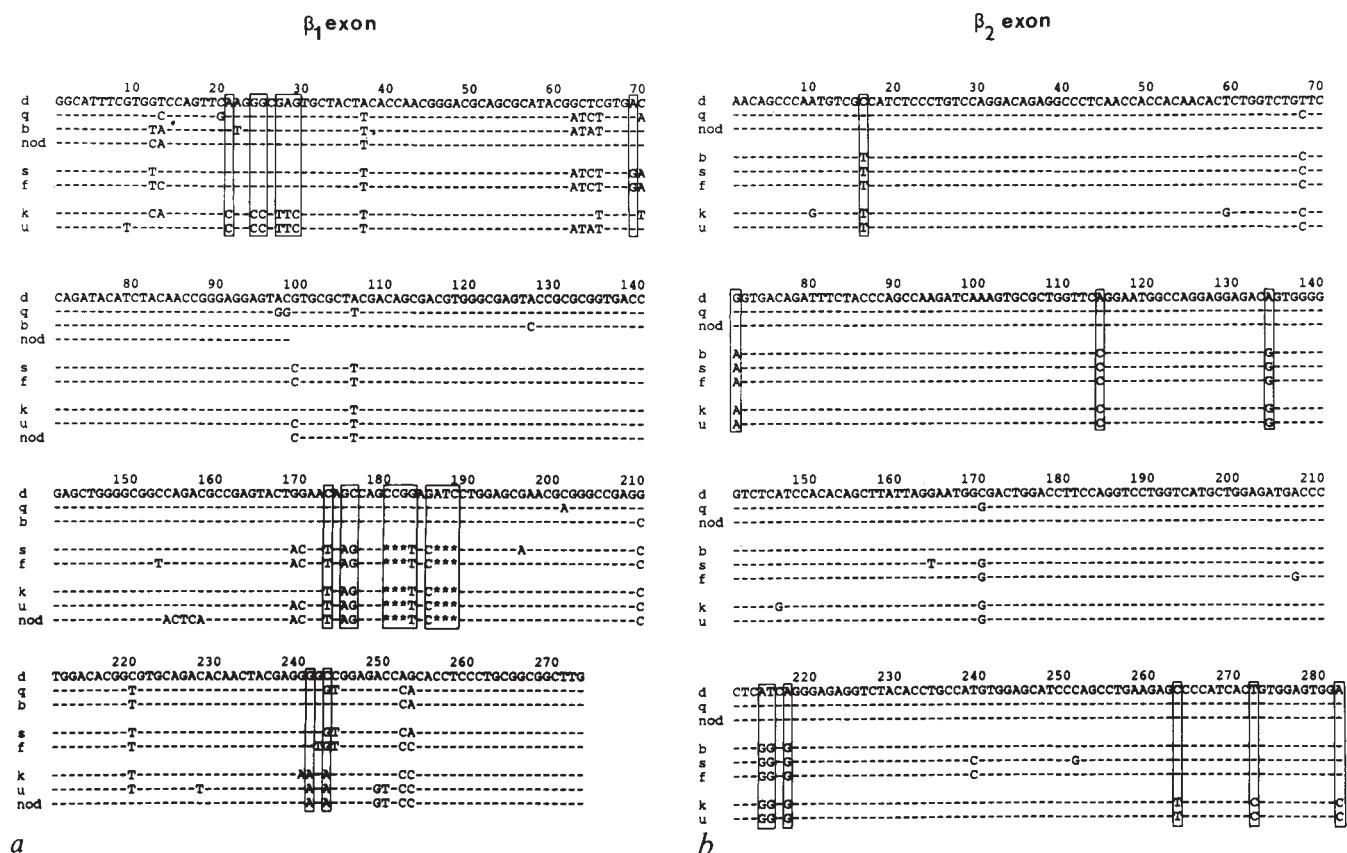


Fig. 2 Nucleotide sequences of β_1 (a) and β_2 (b) exons of eight $I-A_\beta$ alleles^{15,16,20,21}. The sequence of $I-A_\beta^d$ is used as a prototypic sequence: dashes represent identity with $I-A_\beta^d$, polymorphic nucleotides are printed at the appropriate position, and deletions are marked by *. Boxed positions identify sequence polymorphisms which are group-specific. The alleles sequenced had been assigned to evolutionary groups as follows: group 1, d, q, and nod; group 2, b, s, and f; group 3, k and u. Sequences are organized according to the evolutionary origins of the alleles with the exception of the β_1 exons of $I-A_\beta^b$ and $I-A_\beta^{nod}$ which are postulated to have undergone intragenic segmental exchange. The β_1 exon of $I-A_\beta^b$ is grouped with evolutionary group 1 alleles; the 5' portion of the β_1 exon of $I-A_\beta^{nod}$ is grouped with evolutionary group 1 whereas the 3' portion is grouped with evolutionary group 3.

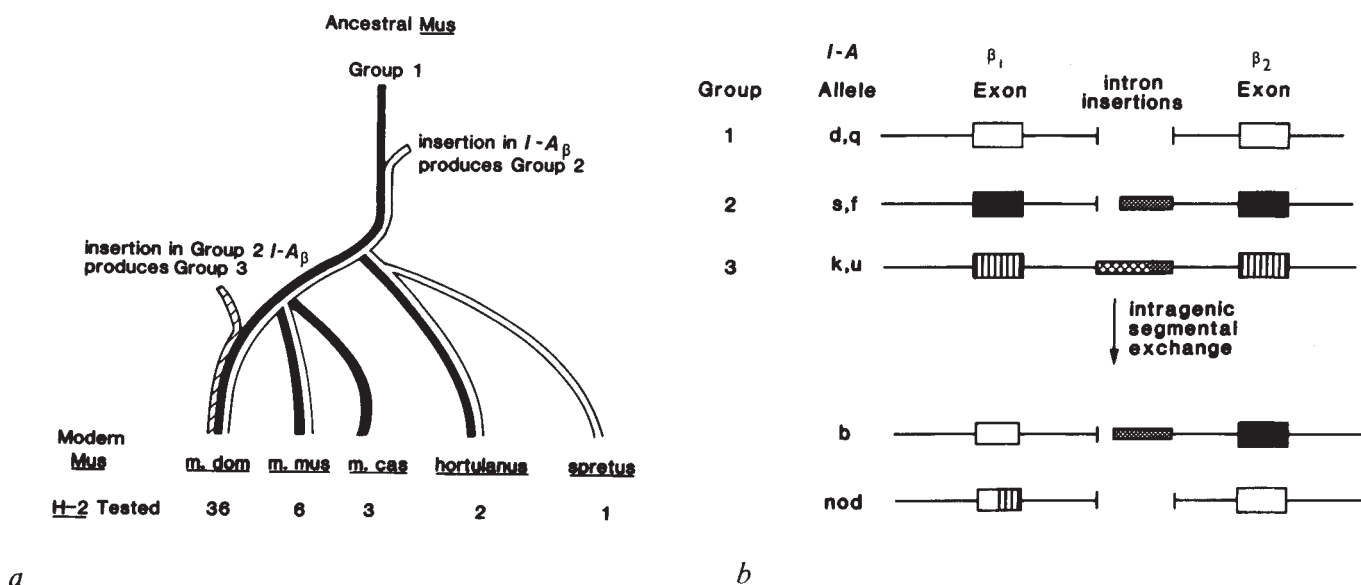


Fig. 3 a, Illustration of the evolutionary origins of the three evolutionary groups of $I-A_\beta$ alleles. The solid line represents evolutionary group 1 which was the earliest form of $I-A_\beta$ in *Mus*. The open line represents evolutionary group 2 which was formed by a retroposon insertion into a group 1 allele before the separation of the five *Mus* species assayed. The hatched line represents evolutionary group 3 which was formed by an additional insertion and/or duplication in a group 2 allele subsequent to the inception of *Mus m. domesticus* (C. C. Lu, C. Zack & E. K. Wakeland, unpublished data). b, A summary of the relationships of the sequence polymorphisms in the β_1 and β_2 exons with the retroposon polymorphisms which occur in the intron between them. Six of the eight $I-A_\beta$ alleles have exon sequence polymorphisms that are associated with the retroposon polymorphism. The remaining two alleles appear to have been produced by intrinsic segmental exchange.

this portion of *I-A_B*. More than 90% of all wild mice tested are *H-2* heterozygotes^{22,23} and the simplest interpretation of the data is that recombination between *I-A_B* alleles in different evolutionary groups is responsible for segmental exchange in the β_1 exon.

We thank Wayne Potts for helpful discussions of the data. This research received support from the NIH.

Received 3 August 1987; accepted 4 February 1988.

- Klein, J. *et al.* *Nature* **291**, 455-460 (1981).
- Mengle-Gaw, L. & McDewitt, H. O. *A. Rev. Immun.* **3**, 367-396 (1985).
- Wedeira, G. & Flavell, R. A. *EMBO J.* **3**, 1221-1225 (1984).
- Klein, J. in *Immunology* **80** (eds Fougereau, M. & Dausset, J.) 239-253 (Academic, New York, 1980).
- Arden, B. & Klein, J. *Proc. natn. Acad. Sci. U.S.A.* **79**, 2342-2346 (1982).
- Klein, J. & Figueroa, F. *CRC Critical Rev. in Immunol.* **6**, 295-386 (1987).
- Sage, R. D. in *The Mouse in Biomedical Research* (eds Foster, H. L., Small, J. D. & Fox, J. G.) **1**, 40-90 (Academic, New York, 1981).
- Ferris, S. D. *et al.* *Genetics* **105**, 681-721 (1983).
- Marshall, J. T. in *Curr. Topics Microbiol. Immunol.* Vol. 127 (eds Potter, M., Nadeau, J. H. & Cancro, M. P.) 12-19 (Springer, Berlin, 1986).
- Wakeland, E. K. & Darby, B. R. *J. Immun.* **131**, 3052-3058 (1983).
- McConnell, T. J., Darby, B. & Wakeland, E. K. *J. Immun.* **136**, 3076-3084 (1986).
- Wakeland, E. K. *et al.* *Immunogenetics* **26**, 115-119 (1987).
- Nei, M. & Li, W. H. *Proc. natn. Acad. Sci. U.S.A.* **76**, 5269-5273 (1979).
- Falus, A. *et al.* *Immunogenetics* **25**, 290-298 (1987).
- Malissen, M., Hunkapillar, T. & Hood, L. *Science* **221**, 750-753 (1983).
- Larhammar, D. *et al.* *Cell* **34**, 179-188 (1983).
- Houck, C. M., Rinehart, F. P. & Schmid, C. W. *J. molec. Biol.* **132**, 289-306 (1979).
- Jagadeeswaran, P., Forget, B. G. & Weissman, S. M. *Cell* **26**, 141-142 (1981).
- Bonhomme, F. in *Curr. Topics Microbiol. Immunol.* Vol. 127 (eds Potter, M., Nadeau, J. H. & Cancro, M. P.) 19-34 (Springer, Berlin, 1986).
- Estess, P. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **83**, 3594-3598 (1986).
- Acha-Orbea, H. & McDewitt, H. O. *Proc. natn. Acad. Sci. U.S.A.* **84**, 2435-2439 (1987).
- Duncan, W. R., Wakeland, E. K. & Klein, J. *Nature* **281**, 603-605 (1979).
- Wakeland, E. K. & Klein, J. in *Current Trends in Histocompatibility* (eds Reisfeld, R. A. & Ferrone, S.) **4**, 289-305 (Plenum, New York, 1981).
- Duncan, W. R. & Klein, J. *Immunogenetics* **10**, 45-65 (1980).
- Winter, W. R. *et al.* *Diabetes* **36**, Suppl. 1, 22A (Abstract) (1987).
- Potts, W. K. *et al.* in *Major histocompatibility genes and their role in immune function* (ed. David, C. S.) (Plenum, New York, in the press).

Independent evolution of structural and coding regions in a *Neurospora* mitochondrial intron

Elizabeth M. Mota & Richard A. Collins*

Department of Botany, University of Toronto, Toronto, Ontario M5S 1A1, Canada

The discovery of intervening sequences (introns) in eukaryotic genes has raised questions about the origin and evolution of these sequences. Hypotheses concerning these topics usually consider the intron as a unit that could be lost or gained over time¹⁻⁶, or as a region within which recombination can occur to facilitate the production of new proteins by exon shuffling⁷. Additional complexities are observed in introns of mitochondrial and chloroplast genes which contain secondary structures required for messenger RNA splicing^{8,9} and open-reading frames encoding proteins^{10,11}. Here we describe differences in the organization of protein-coding sequences in the intron of the mitochondrial *ND1* gene in two closely related species of *Neurospora*. These differences show that intron sequences involved in secondary structure formation and in protein coding can evolve as physically distinct elements. Indeed, the secondary structure elements of the *ND1* intron can contain two different coding sequences located at two different positions within the intron.

The standard laboratory strain of *N. crassa*, 74A, and a representative natural isolate of *N. intermedia*, Varkud-1c, both contain a group I intron at precisely the same position in the *ND1* gene (ref. 12 and Figs 1, 2b; *ND1* probably encodes a

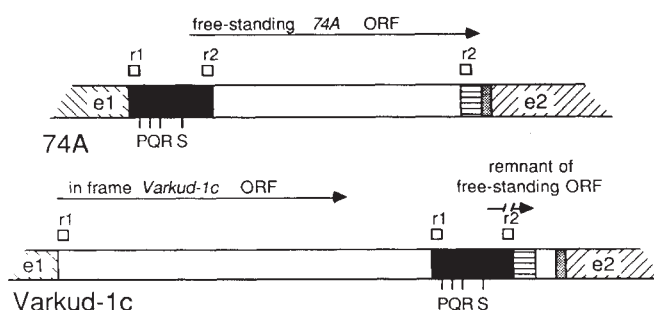


Fig. 1 Organization of the *ND1* gene in different *Neurospora* strains. The map of the standard laboratory strain 74A is derived from the data in ref. 12. Varkud-1c is a *N. intermedia* strain collected from Varkud, India (obtained from the Fungal Genetics Stock Center, strain no. 1823). The DNA sequences of the regions of the genes represented by boxes with matching filled patterns are >97% identical. Unfilled boxes are not detectably related. Exons 1 and 2 are indicated as e1 and e2. P, Q, R and S indicate the conserved sequences involved in formation of the characteristic group I intron secondary structure^{8,9,24}. The positions of direct repeats at or near the sites of gain or loss of intron sequences are marked r1 and r2 (see Fig. 2).

subunit of NADH dehydrogenase¹²). The introns include a sequence of two hundred nucleotides, 97% identical in the two strains, which contains the bases responsible for the core secondary structure characteristic of group I introns^{8,9}. All of the base substitutions in this core structure element maintain the intron secondary structure. This identity in location and similarity in sequence is expected for homologous introns descended from a relatively recent common ancestral *ND1* gene.

In contrast to this conservation of position and secondary structure, the introns in these two strains contain completely dissimilar open-reading frames located at different positions relative to the core structure elements. (Fig. 1). In 74A, the core structure sequences are located near the 5' end of the intron and are followed by a free-standing open reading frame (ORF) initiated at an AUG codon separated from the upstream exon by 183 nucleotides¹². The middle 80% of this ORF is absent from the Varkud-1c intron. Instead, a completely different ORF is present in the Varkud-1c intron and is located in-frame with the upstream exon (Fig. 2a). Unlike the atypical 74A ORF (ref. 12), the Varkud-1c ORF contains homologues of the two conserved twelve amino-acid blocks^{8,13,14} typical of the maturase-like ORFs found in most group I introns.

Although we cannot determine the mechanism of gain or loss of these large intron segments, the sequences of the introns provide some clues. The portion of the Varkud-1c intron that is not present in 74A, including the ORF plus 236 nucleotides of untranslated sequence following the ORF, is bounded by direct repeats of a five-nucleotide sequence. This sequence, labelled r1 in Figs 1 and 2, is present in one copy at the beginning of the 74A intron. The portion of the 74A ORF missing from Varkud-1c is located between direct repeats of a different six-nucleotide sequence that is present in one copy in Varkud-1c (labelled r2 in Fig. 2). This organization of direct repeats would be predicted for transpositional insertion of a maturase-like ORF at the exon/intron junction to create the Varkud-1c arrangement. Alternatively, the lack of this ORF in 74A could have resulted from excision of an in-frame ORF by recombination within direct repeats coincidentally near its ends. Parallel hypotheses could account for the gain or loss of the bulk of the 74A ORF. Loss of part of an intron, apparently by such an excision mechanism, has been observed in *S. pombe*¹⁵. Probable loss of an ORF has also been described in *Kluveromyces* although differences in intron location and other structural features make inferences about the mechanism difficult¹⁶. The possible gain of intron sequences has also been reported¹⁷⁻¹⁹,

* To whom correspondence should be addressed.

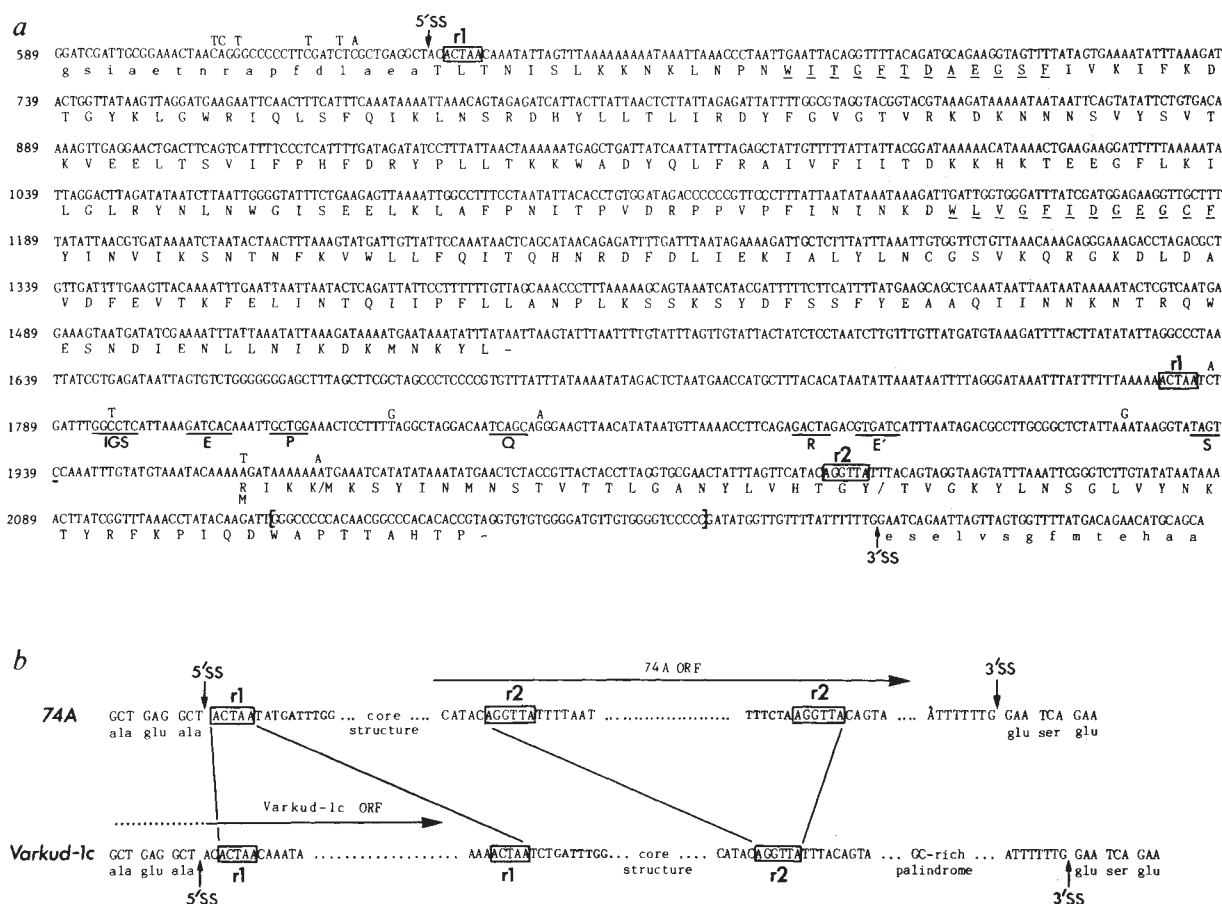


Fig. 2 *a*, DNA sequence of the *ND1* intron in Varkud-1c and translated amino acid sequence of the intron open reading frame. 5'SS and 3'SS indicate the donor and acceptor splice sites respectively, deduced from comparison with the 74A sequence. The sequence begins at base 589 of exon 1 (ref. 12). The sequences of the adjacent 120 nucleotides of exon 1, all of exon 2, and 150 nucleotides of 3' untranslated region are identical in Varkud-1c and 74A (not shown). Exon and intron amino acids are indicated by lower and upper case letters respectively. Homologues of the two conserved blocks of twelve amino acids found in many group I ORFs (refs 8, 13, 14) are underlined. The portions of the conserved nucleotide sequences involved in base-pairing to form the group I secondary structure (IGS, E, P, Q, R, E' and S) (refs 8, 9, 24) are also underlined. In the regions that are conserved between 74A and Varkud-1c, nucleotide sequence differences are indicated above the line. In the portion of the Varkud-1c intron that contains the remnant of a 74A-type ORF, two frame shifts are indicated (/). The positions of direct repeats at or near the sites of gain or loss of intron sequences are marked r1 and r2. A GC-rich palindromic sequence near the 3' end of the Varkud-1c intron is enclosed in brackets. *b*, Comparison of the regions of the 74A and Varkud-1c introns near the sites of gain or loss of intron sequences. The sequences shown for Varkud-1c are also present in two other *N. intermedia* isolates from India (not shown), indicating that this sequence organization is representative of the Indian population.

Methods. DNA sequences of mtDNA restriction fragments cloned into m13mp18 or 19 were determined by the dideoxy method²⁵ with 7-deaza dGTP substituted for dGTP²⁶. In some reactions dITP²⁷ and Sequenase²⁸ (United States Biochemicals) were used.

in at least one of these cases short direct repeats flank the additional sequence^{17,18}.

However, neither of the above hypotheses precisely explains ORF gain or loss in the *ND1* intron. The sequences adjacent to the direct repeats (r1 and r2) are not exactly those predicted to result from homologous recombination within repeats of the existing sequence or from typical target site duplication induced by insertion of a mobile element. In each intron a small number of extra nucleotides are present adjacent to one of the repeats (Fig. 2*b*). Also, a cluster of six base substitutions is present near the 3' end of the upstream exon (Fig. 2*a*) that may be related to the gain or loss of the in-frame ORF. Clusters of substitutions at the ends of exons that are otherwise highly conserved have also been found adjacent to sites of gain or loss of entire optional introns in yeast¹⁴ and *Neurospora* mtDNA (R. Collins, unpublished data). Hensgens *et al.*¹⁴ have suggested that these clusters may indicate the involvement of larger regions of the chromosome in intron insertion or deletion events, perhaps implicating gene conversion, non-homologous recombination or error-prone repair. Additional changes may also have occurred that have

obscured the sequences resulting directly from the insertion or deletion events involving the ORFs. Indeed, some sequence polymorphisms (at least six base substitutions) are scattered throughout even the conserved portions of the *ND1* introns. Also, a 56-nucleotide GC-rich palindromic sequence, somewhat reminiscent of the optional GC-rich clusters found in other fungal mitochondrial DNAs²⁰⁻²³, is present near the 3' end of the Varkud-1c intron, but absent from the 74A intron (Fig. 2*a*).

From an evolutionary perspective, the discovery that two different protein-coding regions can exist within the same intron structure indicates that at least some introns should be viewed as a composite of two independent genetic elements. One contains the structural information necessary for splicing pre-mRNA from the host gene containing the intron; the other can be a separately evolving gene within a gene. Complete genes within mitochondrial introns are not uncommon²⁹⁻³¹. Moreover, the discovery of genes within introns in the *Drosophila* nuclear genes *Gart*³² and *dunce*³³ suggests that complex organization of intron genetic information is not limited to organelle genomes.

We thank Barry Saville and Andrea Nicholson for cloning

the *ND1* gene from Varkud-1c mitochondrial DNA. This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada and the University of Toronto.

Received 11 December 1987; accepted 19 February 1988.

1. Sharp, P. A. *Cell* **42**, 397-400 (1985).
2. Rogers, J. *Nature* **315**, 458-459 (1985).
3. Doolittle, W. F. *Nature* **272**, 581-582 (1978).
4. Darnell, J. E. *Science* **202**, 1257-1260 (1978).
5. Cavalier-Smith, T. *J. cell Sci.* **34**, 247-278 (1978).
6. Crick, F. *Science* **204**, 264-271 (1979).
7. Gilbert, W. *Nature* **271**, 501 (1978).
8. Michel, F., Jacquier, A. & Dujon, B. *Biochimie* **64**, 867-881 (1982).
9. Davies, R. W., Waring, R. B., Ray, J. A., Brown, T. A. & Scazzocchio, C. *Nature* **300**, 719-724 (1982).
10. Borst, P. & Grivell, L. A. *Nature* **289**, 439-440 (1981).
11. Kotylak, Z., Lazowska, J. & Slonimski, P. P. in *Achievements and Perspectives of Mitochondrial Research*. Vol. 2. (eds E. Quagliariello *et al.*) 1-20 (Elsevier, Amsterdam, 1985).
12. Burger, G. & Werner, S. *J. molec. Biol.* **185**, 231-242 (1985).
13. Waring, R. B., Brown, T. A., Ray, J. A., Scazzocchio, C. & Davies, R. W. *Proc. natn. Acad. Sci. U.S.A.* **79**, 6332-6336 (1982).
14. Hensgens, L. A., Bonen, L., deHaan, M., van der Horst, G. & Grivell, L. A. *Cell* **32**, 379-389 (1983).
15. Merlos-Lange, A. M., Kanbay, F., Zimmer, M. & Wolf, K. *Molec. gen. Genet.* **206**, 273-278 (1987).
16. Dujon, B., Colleaux, L., Jacquier, A., Michel, F. & Montelliet, C. in *Extrachromosomal Elements in Lower Eukaryotes* (eds R. B. Wickner *et al.*) 5-28 (Plenum, New York, 1986).
17. Waring, R. B., Brown, T. A., Ray, J. A., Scazzocchio, C. & Davies, R. W. *EMBO J.* **3**, 2121-2128 (1984).
18. Lang, B. F. *EMBO J.* **3**, 2129-2136 (1984).
19. Michel, F. & Cummings, D. J. *Curr. Genet.* **10**, 69-79 (1985).
20. de Samaroczy, M. & Bernardi, B. *Gene* **41**, 1-22 (1986).
21. Coruzzi, G., Bonitz, S. G., Thalenfeld, B. E. & Tzagoloff, A. *J. biol. Chem.* **256**, 12780-12787 (1981).
22. Michel, F. *Curr. Genet.* **8**, 307-317 (1984).
23. Seraphin, B., Simon, M. & Faye, G. *Nucleic Acids Res.* **13**, 3005-3014 (1985).
24. Burke, J. M. *et al. Nucleic Acids Res.* **15**, 7217-7221 (1987).
25. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
26. Mizusawa, S., Nishimura, S. & Seela, F. *Nucleic Acids Res.* **14**, 1319-1324 (1986).
27. Mills, D. R. & Kramer, F. R. *Proc. natn. Acad. Sci. U.S.A.* **76**, 2232-2235 (1979).
28. Tabor, S. & Richardson, C. C. *Proc. natn. Acad. Sci. U.S.A.* **84**, 4767-4771 (1987).
29. Burke, J. & RajBhandary, U. L. *Cell* **31**, 509-520 (1983).
30. Jacquier, A. & Dujon, B. *Cell* **41**, 383-394 (1983).
31. Macreadie, I. G., Scott, M. R., Zinn, A. R. & Butow, R. A. *Cell* **41**, 395-402 (1983).
32. Henikoff, S., Keene, M. A., Fachtel, K. & Fristrom, J. W. *Cell* **44**, 33-42 (1986).
33. Chen, C., Malone, T., Beckendorf, S. K. & Davis, R. L. *Nature* **329**, 721-725 (1987).

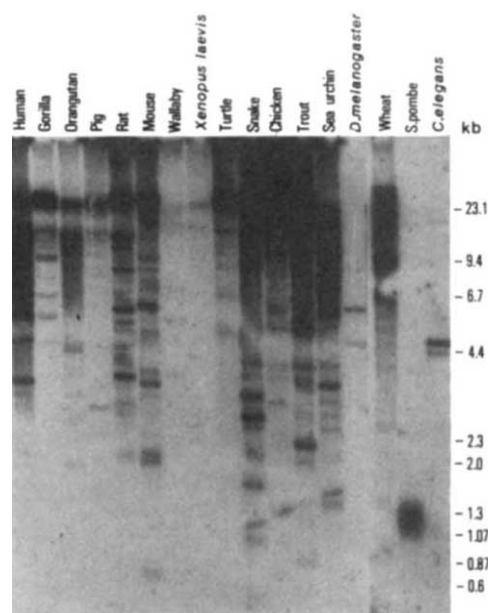


Fig. 1 Cross-hybridization of SPT16 to DNA from a variety of species. Total DNA from the species indicated was digested with an excess of the restriction endonuclease *Eco*RI (3 units per μ g, 37 °C, overnight). The amount of DNA loaded on to a 0.8% agarose gel was $\sim 10 \mu$ g, except for the *D. melanogaster* and *S. pombe* samples, of which 1 μ g and 300 ng were loaded respectively. After electrophoresis the gel was blotted to Hybond and hybridized to 32 P-labelled *Bam*HI/*Hind*III telomere fragment purified from pucSPT16. Hybridization was overnight at 68 °C in 7% SDS, 0.5 M Na_2HPO_4 , pH 7.2. Filters were washed 3 times at 68 °C in $0.1 \times \text{SSC}$, 0.1% sodium pyrophosphate and 0.1% SDS. Size markers (kb) are indicated.

Telomeric repeat from *T. thermophila* cross hybridizes with human telomeres

Robin C. Allshire*, John R. Gosden*, Sally H. Cross*, Gwen Cranston*, Derek Rout*, Neal Sugawara†, Jack W. Szostak†, Peter A. Fantes‡ & Nicholas D. Hastie*

* MRC Clinical and Population Cytogenetics Unit, Western General Hospital, Edinburgh EH4 2XU, UK
† Massachusetts General Hospital, Boston, Massachusetts 02114, USA

‡ Department of Zoology, University of Edinburgh, Edinburgh EH9 3JT, UK

The ends (telomeres) of eukaryotic chromosomes must have special features to ensure their stability and complete replication. Studies in yeast¹⁻³, protozoa⁴⁻⁶, slime moulds^{7,8} and flagellates^{8,10} show that telomeres are tandem repeats of simple sequences that have a G-rich and a C-rich strand. Mammalian telomeres have yet to be isolated and characterized, although a DNA fragment within 20 kilobases of the telomeres of the short arms of the human sex chromosomes has been isolated¹¹. Recently we showed that a chromosome from the fission yeast *Schizosaccharomyces pombe* could, in some cases, replicate as an autonomous mini-chromosome in mouse cells¹². By extrapolation from other systems^{1,3,13,14}, we reasoned that mouse telomeres could be added to the *S. pombe* chromosome ends in the mouse cells. On setting out to test this hypothesis we found to our surprise that the telomeric probe used (containing both the *S. pombe* and *Tetrahymena thermophila*

repeats) hybridized to a series of discrete fragments in normal mouse DNA and DNA from a wide range of eukaryotes. We show here that the sequences hybridizing to this probe are located at the telomeres of most, if not all, human chromosomes and are similar to the *Tetrahymena* telomeric-repeat component of the probe.

The telomeric tandem repeat sequence from *S. pombe* has recently been cloned³. Owing to the cloning strategy, the resultant plasmid pSPT16 contained ~ 100 base pairs (bp) of the *T. thermophila* telomeric repeat, $(\text{C}_4\text{A}_2)_N$ and ~ 300 bp of the *S. pombe* telomeric repeat, $(\text{C}_{1-6}\text{G}_{0-1}\text{T}_{0-1}\text{GTA}_{1-2})_N$.

Figure 1 shows the hybridization of this probe to a variety of eukaryotic DNAs. In a number of cases there is a variable smear in addition to a series of discrete bands. This is a provocative finding because individual telomeres in several organisms show considerable size heterogeneity¹⁻¹¹. The total signal seen in these organisms is variable and may be proportional to the number of linkage groups, although the signal in some species (for example, *Xenopus laevis* and the wallaby) is relatively weak.

A strong signal has also been seen in other organisms (tobacco, petunia and crab, data not shown), but we have not detected any hybridization of the probe to DNA extracted from *S. cerevisiae* or *E. coli* (data not shown). These hybridizations were all washed at high stringency ($0.1 \times \text{SSC}$; 68 °C). There must therefore be a considerable degree of similarity between the probe and the sequences detected.

Next we tested whether these cross-hybridizing sequences could be located at, or near, human telomeres. Firstly, we looked for differences in the pattern of hybridization observed between sperm and somatic DNA from the same individual. Cooke and Smith¹⁵ showed that telomeres of the human sex chromosome were several kilobase pairs (kb) longer in sperm DNA relative to somatic DNA. Figure 2 shows the digestion pattern seen when the SPT16 probe was hybridized to sperm and somatic

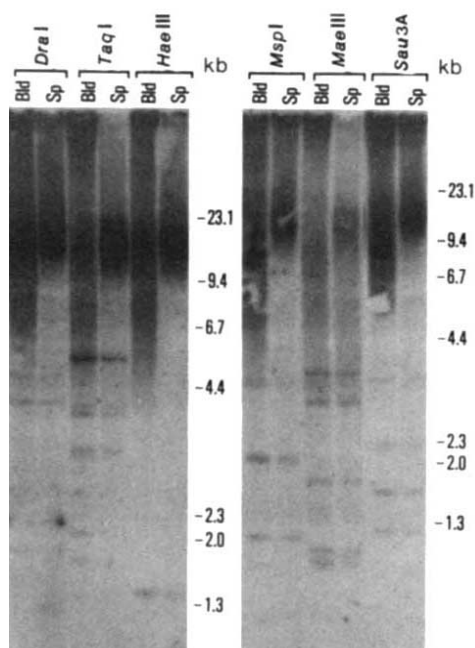


Fig. 2 Comparison of germline and somatic DNA samples from a single individual. Blood (Bld) and sperm (Sp) DNAs ($\sim 10 \mu\text{g}$) were digested with the restriction enzymes indicated and analysed by blot-hybridization (conditions and probe as in Fig. 1). Size markers (kb) are indicated.

DNA from the same individual using a variety of restriction enzymes. The smear in all cases is on average of higher molecular weight and more homogeneous in sperm DNA than in somatic DNA, indicating that the probe could be recognizing human telomeres. A very similar pattern is seen with a range of restriction enzymes. This indicates that the region detected by SPT16 is devoid of these restriction enzyme sites and probably consists of a simple repetitive sequence. This region ranges from 5–20 kb in blood and 10–20 kb in sperm.

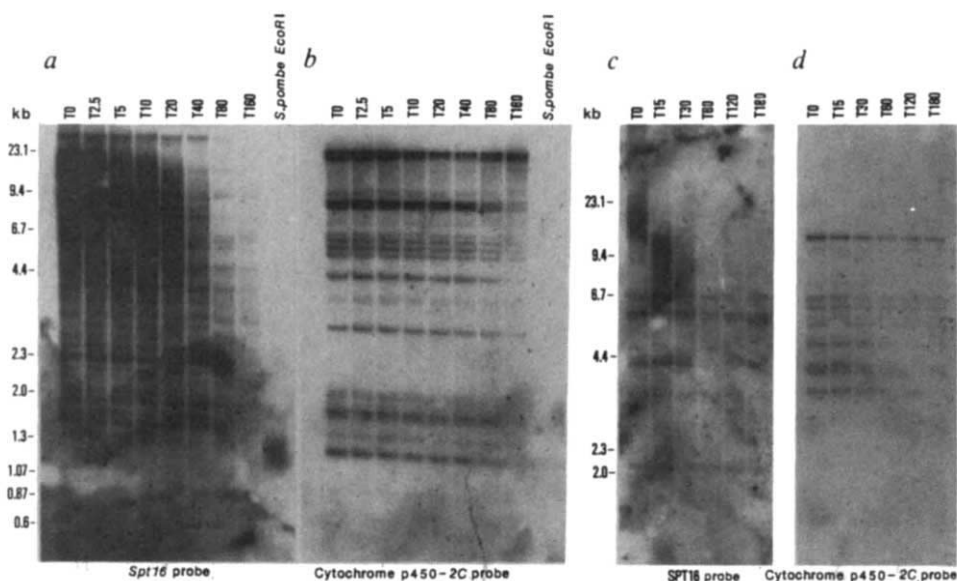
The definitive test of telomere location is sensitivity to digestion by the exonuclease *Bal31*, because the telomeres of lower eukaryotes^{1,3,5,6,9,10} and the human sex chromosomes¹⁵ are all accessible to digestion by this enzyme. Figure 3a and c shows hybridization of the SPT16 probe to human placenta or sperm DNA that has been digested for increasing periods of time with

Bal31 and then with the restriction endonucleases *MspI* and *TaqI* respectively. It is noticeable that the diffuse smear of hybridization ranging in size from 5 to 20 kb in placenta DNA (Fig. 3a) has completely disappeared after 40–80 min incubation. By contrast, the size of most of the discrete bands remains undiminished for up to 160 min of *Bal31* treatment. Panel 3b shows rehybridization of the same filter to a probe for the cytochrome P450-C2C gene family which resides at chromosome 10q23–24 (ref. 16), a long distance from the telomere. There is no decrease in size of the P450-C2C-hybridizing bands during the time course.

Thus the heterogeneous smear that is sensitive to *Bal31* digestion must represent sequences close to the telomere. Using the telomeric probe for the human sex chromosomes¹¹ we have estimated that the rate of removal of nucleotides by *Bal31* in Fig. 3a and b was ~ 3.8 kb per hour (data not shown). As the entire smear has almost gone within 40 min, the sequences detected must be within 2.5 kb of the end of the chromosomes in placenta. In Fig. 3c and d it can be seen that the large 10–20 kb smear of hybridization disappears within 60 min of *Bal31* digestion, but that the cytochrome P450-C2C signal remains intact throughout. Clearly the more homogeneous smear in sperm DNA is also located close to the telomere. In Fig. 3 the discrete fragments hybridizing to the telomere probe appear to be no more sensitive to *Bal31* digestion than the cytochrome P450-C2C genes. These discrete fragments are likely to reside at internal chromosomal sites. Whether such internal sites are at a telomeric location or not remains to be seen (see below). The prominence of these discrete bands in Fig. 3a compared with Fig. 2 is a result of the different hybridization conditions used (see figure legends).

If the SPT16 probe is hybridizing to telomeres, this should also be revealed by *in situ* hybridization to metaphase chromosome spreads. Figure 4 shows a representation of the grain distributions scored after hybridization of the telomere probe *in situ* to metaphase spreads of three different individuals (depicted by different symbols). Fifty-four per cent of the grains were located in the telomeric bands, compared with 5–15% for various single-copy probes mapped in our laboratory^{16–18}. There are two other potentially interesting features revealed by these data. Firstly, there appears to be variation among the three individuals in the range of chromosomes to which the probe hybridizes strongly. For example, in male 2, 14.3% (8% of total grains scored) of telomeric grains (54% of total grains scored) are at 13qter, whereas in male 1, only 4.8% of grains are found at this telomere, and in female 1 there are none. In contrast, 9.5% of

Fig. 3 *Bal31* sensitivity of the diffuse smear homologous to SPT16 in human DNA. Human male placental DNA (a and b) or sperm DNA (c and d) was digested with *Bal31* for the times (T) indicated. The samples were then digested with *MspI* (a and b) or *TaqI* (c and d), separated by electrophoresis in agarose, blotted to nitrocellulose filters and hybridized with ³²P-labelled *Bam*HI *Hind*III fragment from pucSPT16 (a and c). Hybridization conditions for nitrocellulose were as previously described¹². Filters were washed as described in Fig. 1. After autoradiographic exposure, the filters were washed in boiling distilled water and were then rehybridized with ³²P-labelled probe for the P450-C2C gene family¹⁶; the resulting autoradiographs are shown in b and d. Size markers (kb) are indicated.



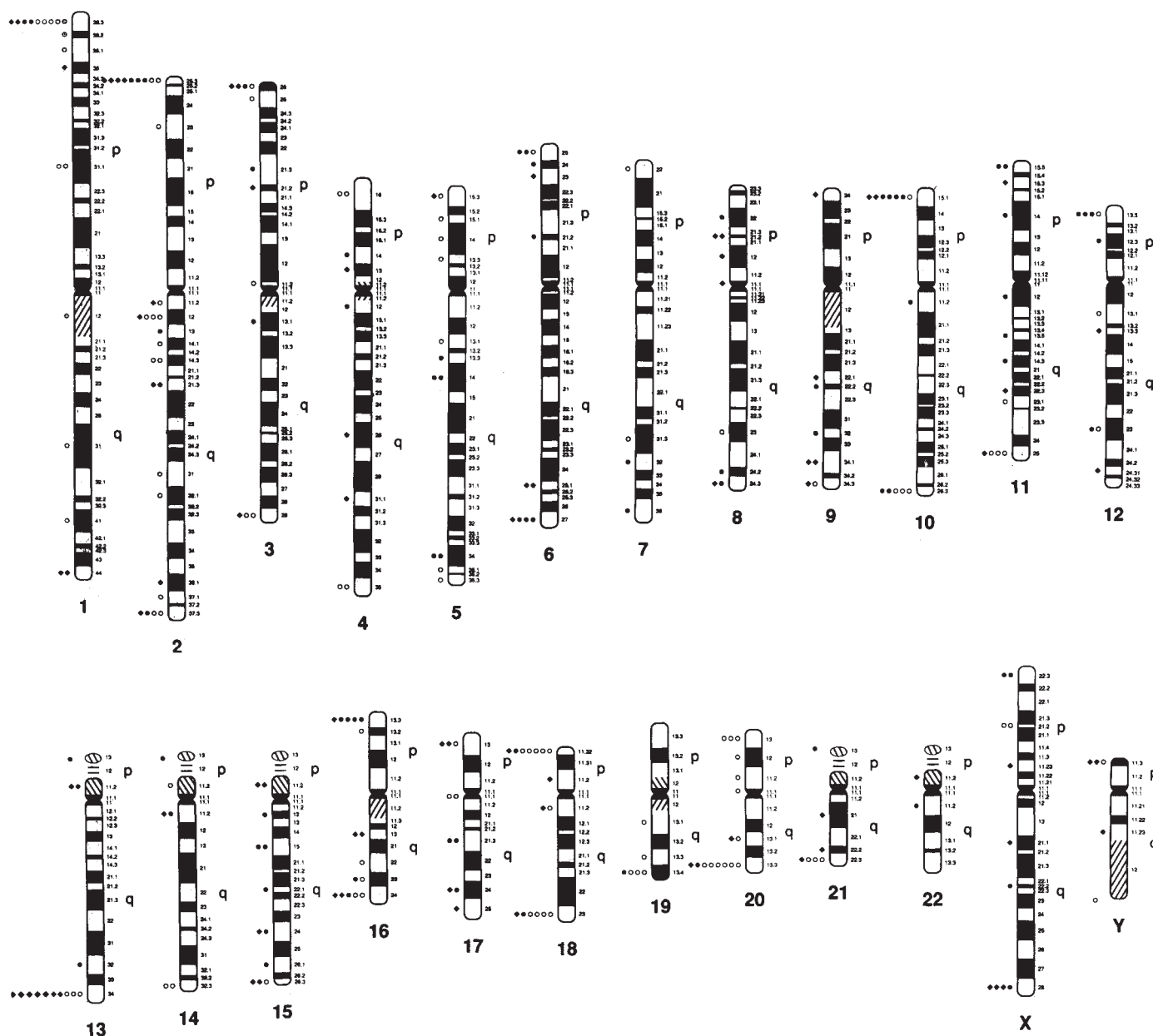


Fig. 4 *In situ* hybridization of SPT16 to metaphase spreads of human chromosomes. The positions of grains observed in 37 metaphase spreads prepared from two male individuals (male 1 and male 2 represented by ○ and ◆) and one female (female 1 represented by ●) are shown. A total of 274 grains were counted, 149 of which are located in the most telomeric bands. Metaphase spreads and ³H-labelled SPT16 probe (specific activity ~2 × 10⁸ d.p.m. μg⁻¹) were prepared as described previously^{12,17}. Hybridization and autoradiography were as described¹². The final washes (4 × 5 min) were at 40 °C in 1 × SSC.

the telomeric grains seen in male 1 map to the 20qter, male 2 and female 1 having only 2% and 2.7% of their telomeric grains at this position. If this is confirmed later it would imply either that individual telomeres are undergoing cycles of expansion and contraction (as demonstrated in *T. thermophila*¹⁹ and *T. Brucei*^{20,21}), or that there is telomere exchange between non-homologous chromosomes. Secondly, there is one internal, non-telomeric site of hybridization located at 2q11-14. This is particularly interesting because chromosome 2 is thought to have been formed by fusion of two ancestral acrocentric ape chromosomes²²⁻²⁴. It has been suggested that this was a centromeric fusion resulting in the loss of the short arm telomeres of the two chromosomes involved (ref. 22), but the alternative hypothesis of the two chromosomes being joined by telomere fusion^{23,24}, is supported by hybridization of the telomere probe at bands 2q11-2q14. It is also of interest that 2q11 and 2q13

are folate-sensitive fragile sites²⁵. This fragility could possibly arise from the residual telomeric sequences.

We conclude that the SPT16 probe recognizes telomeres of humans and, in all likelihood, of other higher eukaryotes. Under our hybridization conditions, the telomeres of *T. thermophila*, *S. cerevisiae* and *S. pombe* failed to cross-hybridize, so the signal we observe probably reflects a real sequence identity, rather than an overall G-richness of the probe (Fig. 5). Furthermore, the use of other probes containing only the *S. pombe* or *T. thermophila* telomeric sequences has established that the signal detected in human and mouse DNA is due to the *T. thermophila* (C₄A₂)_N component of the SPT16 probe (Fig. 5). The *S. pombe* telomeric repeat alone does not hybridize to total human, mouse or *S. cerevisiae* DNA or to *T. thermophila* ribosomal DNA telomeres (Fig. 5c). Therefore human telomeres have a sequence similar to those of *T. thermophila*.

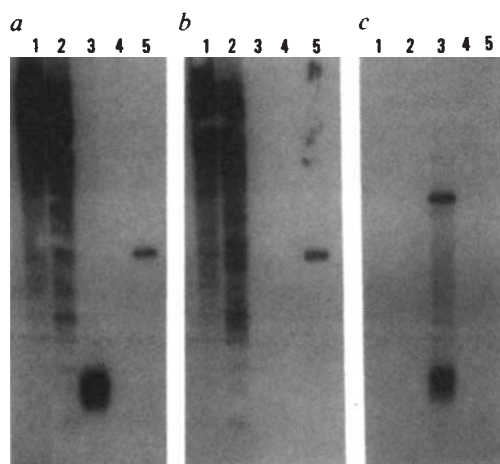


Fig. 5 Hybridization with *S. pombe* and *T. thermophila* telomeric probes. *a*, BamHI/HindIII fragment from pucSPT16 containing both *T. thermophila* (100 bp) and *S. pombe* (300 bp) telomeric repeats; *b*, BamHI/XhoI fragment from YAC4 containing the *T. thermophila* telomeric repeat (300 bp); *c*, SacI/HindIII fragment from pNSU68 containing the *S. pombe* telomeric repeat (200 bp). Lanes showing total DNA from (1) human placenta, 10 µg; (2) mouse C₁₂₇ cells, 10 µg; (3) *S. pombe* ED628, 300ng (4) *S. cerevisiae* DC5, 300ng digested with EcoRI and loaded on three identical 0.8% agarose gels. Lanes(5), a purified 3.5-kb telomeric BglII fragment (~80 pg) from *T. thermophila* rDNA. After electrophoresis, the gels were blotted to Hybond and hybridized overnight at 68 °C in 7% SDS, 0.5 M Na₂HPO₄, pH 7.2, to the ³²P-labelled purified telomeric probes. Filters in *a* and *b* were washed as in Fig. 1. Filter in *c* was washed in 1 × SSC at 68 °C.

All known telomeres have one protruding strand which is G-rich. This must also be true of human telomeres, shown here to cross-hybridize to the (C₄A₂ · T₂G₄)_N *T. thermophila* telomeric repeat. Recently, it has been shown²⁶ that oligonucleotides of the *T. thermophila* telomere G-strand form novel intramolecular structures. Also, only oligonucleotides of the G-rich strand from the telomeres of five different organisms can act as elongation primers for the telomerase present in *T. thermophila* cell extracts²⁷. Given that the *T. thermophila* telomeric repeat cross-hybridizes to human telomeres, we would also expect these other properties to be similar. As *T. thermophila* and humans are separated by a large evolutionary distance, it is possible that only a limited number of G-rich sequences can function as telomeres in eukaryotes.

Our results with this (C₄A₂)_N probe point to a direct isolation of the telomeres of individual human chromosomes. Beyond defining their structural properties, cloned human telomeres could be useful in the construction of synthetic functional chromosomes, in the study of possible interactions between homologous and non-homologous chromosomes, to characterize chromosome rearrangements involving telomeres, and to define the genetic limits of each chromosome. In this last regard, it is of particular interest that the Huntington's chorea gene is thought to reside very close to the telomere of the short arm of chromosome 4 (ref. 28). No flanking markers on the distal side of this mutation have been identified. Thus the telomere of the short arm of chromosome 4 could allow a better definition of the locus and facilitate the isolation of the gene itself.

We thank Richard Meehan, Bob Hill, Adrian Bird, David Finnegan, Jane Kinnard, John Sulston, Anne Ramage and Steve Smith for various DNA samples; Howard Cooke and members of our laboratories for their invaluable advice and discussion, and especially Richard Buckland for pointing out the history of human chromosome 2; N. Davidson, S. Bruce and D. Stewart for photographic work; and Professor H. J. Evans for his encouragement and support. S.H.C. is supported by SERC studentship.

Received 27 January; accepted 7 March 1988.

1. Shampay, J., Szostak, J. W. & Blackburn, E. H. *Nature* **310**, 154-157.
2. Walmsley, R. W., Chan, C. S. M., Tye, B. K. & Petes, T. D. *Nature* **310**, 157-160 (1984).
3. Sugawara, N. & Szostak, J. W. *Yeast* **2** (suppl.) 373 (1986).
4. Blackburn, E. H. & Gall, J. G. *J. molec. Biol.* **120**, 33-53 (1978).
5. Klobutcher, L. A., Swanton, M. A., Donini, P. & Prescott, D. M. *Proc. natn. Acad. Sci. U.S.A.* **78**, 3015-3019 (1981).
6. Ponzi, M., Pace, T., Dore, E. & Frontali, C. *EMBO J.* **4**, 2991-2995 (1985).
7. Johnson, E. M. *Cell* **22**, 875-886 (1980).
8. Emery, H. S. & Weiner, A. M. *Cell* **26**, 411-419 (1981).
9. Blackburn, E. H. & Challoner, P. B. *Cell* **36**, 447-457 (1984).
10. Van der Ploeg, L. H. T., Lin, A. Y. C. & Borst, P. *Cell* **36**, 459-468 (1984).
11. Cooke, H. J., Brown, W. R. A. & Rappold, G. A. *Nature* **317**, 687-692 (1985).
12. Allshire, R. C. *et al. Cell* **50**, 391-403 (1987).
13. Pluta, A. F., Dani, G. M., Spear, B. B. & Zakian, V. A. *Proc. Acad. Sci. U.S.A.* **78**, 1475-1479 (1984).
14. Perrot, M., Barreau, C. & Bégueret, J. *Molec. cell. Biol.* **7**, 1725-1730 (1987).
15. Cooke, H. J. & Smith, B. A. *Cold Spring Harbor Symp. Quant. Biol.* **51**, 213-219 (1986).
16. Meehan, R. R. *et al. Am. J. hum. Genet.* **42**, 26-37 (1988).
17. Gosden, J. R., Middleton, P. G., Rout, D. & De Angelis, C. *Cytogenet. Cell Genet.* **43**, 150-153 (1986).
18. Liou, G. I. *et al. Somat. Cell. molec. Genet.* **13**, 315-323 (1987).
19. Larson, D. D., Spangler, E. A. & Blackburn, E. H. *Cell* **50**, 477-483 (1987).
20. Bernards, A., Michels, P. A. M., Lincke, C. R. & Borst, P. *Nature* **303**, 592-547 (1983).
21. Pays, E., Laurent, M., De Linte, K., Van Meirvenne, N. & Steinert, M. *Nucleic Acids Res.* **11**, 8137-8147 (1983).
22. De Grouchy, J., Turleau, C., Finaz, C. & Roubin, M. *Chromosome Variation in Human Evolution* (ed. Boyce, A. J.) 17-38 (Taylor & Francis, London, 1975).
23. Lejeune, J., Dutrillaux, B., Rethoré, M. O. & Prieur, M. *Chromosoma* **43**, 423-444 (1973).
24. Yunis, J. J. & Prakash, O. *Science* **215**, 1525-1530 (1982).
25. Sutherland, G. R. & Hecht, F. *Oxford Monographs on Medical genetics* **13**, (Oxford University Press, 1985).
26. Henderson, E. *et al. Cell* **51**, 899-908 (1987).
27. Greider, C. W. & Blackburn, E. H. *Cell* **51**, 887-898 (1987).
28. Gilliam, T. C. *et al. Cell* **50**, 565-571 (1987).

Liquid-like movements in crystalline insulin

D. L. D. Caspar, J. Clarage, D. M. Salunke & M. Clarage

Rosenstiel Basic Medical Sciences Research Center, and Department of Physics, Brandeis University, Waltham, Massachusetts 02254, USA

Diffuse X-ray scattering from protein crystals provides information about molecular flexibility and packing irregularities¹⁻⁴. Here we analyse diffraction patterns from insulin crystals that show two types of scattering related to disorder: very diffuse, liquid-like diffraction, and haloes around the Bragg reflections. The haloes are due to coupled displacements of neighbouring molecules in the lattice, and the very diffuse scattering results from variations in atomic positions that are only locally correlated within each molecule. The measured intensity was digitally separated into three components: the Bragg reflections and associated haloes; the water and Compton scattering; and the scattering attributed to internal protein movements. We extend methods used to analyse disorder in membrane structures⁵⁻⁷ to simulate the diffuse scattering from crystalline insulin in terms of (1) the Patterson (autocorrelation) function of the ideal, ordered crystal structure, (2) the root-mean-square (r.m.s.) amplitude of the atomic movements, and (3) the mean distance over which these displacements are coupled. Movements of the atoms within the molecules, with r.m.s. amplitudes of 0.4-0.45 Å, appear to be coupled over a range of ~6 Å, as in a liquid. These locally coupled movements account for most of the disorder in the crystal. Also, the protein molecules, as a whole, jiggle in the lattice with r.m.s. amplitudes of ~0.25 Å that appear to be significantly correlated only between nearest neighbours.

The scattering arising from structural variations will be designated 'variational', and that attributable to the average structure (the norm) will be called 'normal'. These terms are preferred, rather than names such as 'diffuse' and 'crystalline', to describe the two types of diffraction because variations relative to some ideally ordered molecular structure are not confined to crystals and some perturbations are periodic, which produce sharp, rather than diffuse, variational scattering (for example, ref. 8).

The only variations considered here are in the positions of atoms in chemically identical molecules. Mean-square displacements of the atoms in the unit cell of a crystal, relative to their ideal, most probable positions, are measured by the Debye-Waller temperature factors⁹⁻¹². No information about coupling among the atomic movements is contained in the normal Bragg diffraction data used to calculate the temperature factors. The variational scattering contains information about both the amplitudes of the displacements and the correlations in the molecular movements^{1-8,10-12}.

We recorded X-ray scattering from a stationary 2Zn pig insulin crystal¹³ (space group R3, hexagonal unit cell $a = 82.5 \text{ \AA}$, $c = 33.9 \text{ \AA}$) taking great care to minimize the extraneous background (Fig. 1a). Comparisons of short and long exposures indicate that the mean intensity of the variational scattering from the well-ordered, crystalline protein is about a thousandth that of strong Bragg reflections; thus, this diffuse scattering is quite weak compared to the background of conventional X-ray cameras. Brief exposures, taken before and after recording the variational scattering, established that radiation damage was negligible.

Variational scattering due to fluctuations that are only locally correlated within each molecule must vary smoothly as a function of scattering angle, with no harmonics in the intensity distribution of wavelength shorter than the reciprocal of the molecular dimensions. Haloes about the Bragg reflections, due to correlated displacements of the molecules in the lattice, show higher spatial frequency modulation. We measured the very diffuse scattering component from the digitized film data (Fig. 1a) by smoothly interpolating between local intensity minima. Subtraction of this component from the film data gave a zero background (within the noise level of the film) between layer lines in the c^* direction (Fig. 1b). The widths of the haloes about the Bragg reflections in Fig. 1b, as defined by this partitioning of the data, are therefore less than $c^* (= 0.029 \text{ \AA}^{-1})$.

Compton and water scattering, which constitute part of the diffuse scattering, should be spherically symmetrical. Beyond about $0.5 \text{ \AA}^{-1}$ spacing, where the Bragg reflections begin to fade out, the diffuse scattering from the insulin crystal is nearly spherically symmetrical. The Compton scattering was scaled to the circularly averaged high-angle data according to the ratio of the inelastic to elastic scattering cross-sections of the atoms¹⁴. Extraneous scattering from water and the usual capillary container was avoided by mounting the crystal on a thin glass rod in a moist air stream at slightly below saturation, which eliminated adhering liquid without affecting crystal order¹⁵. Water in the crystal was assumed to scatter like the pure liquid¹⁶. This scattering was scaled to the high-angle data (corrected for Compton scattering) in proportion to the number of water electrons in the insulin unit cell^{13,17}. A wedge of the estimated water and Compton scattering is shown superimposed in Fig. 1b. Subtraction of this circularly symmetrical component from the total diffuse scattering gave the component attributed to short-range variations in the protein structure shown in Fig. 1c. Significant features in this surprisingly asymmetrical intensity distribution can be accounted for in terms of the structure and internal flexibility of the symmetrically arranged insulin molecules in the crystal.

Figure 2a shows the α -carbon backbones of the six insulin molecules in the rhombohedral unit cell¹³, oriented as viewed along the X-ray beam. This direction of view, which is nearly perpendicular to the threefold axis and 10° to the left of a quasi-twofold axis of the hexamer, minimizes the symmetry of the projected molecular structure and, therefore, enhances the information content in the diffraction photograph of the stationary crystal. The intensities of the Bragg reflections are determined by the normal unit-cell scattering function $|F_N(\mathbf{R})|^2$ plotted for the measured crystal orientation in Fig. 2b together with the reciprocal lattice points calculated¹⁸ to be in reflecting position. $\mathbf{R}$ is the reciprocal space scattering vector of length

$2 \sin \theta / \lambda$. $|F_N(\mathbf{R})|^2$ was evaluated from protein data-bank coordinates^{19,20} as the product of the square of the transform $|F_0(\mathbf{R})|^2$ of the refined ideal atomic structure of one unit cell and the Debye-Waller function, $e^{-(2\pi\delta_N\mathbf{R})^2}$. The average r.m.s. atomic displacement amplitude $\delta_N = \langle \delta_N^2 \rangle^{1/2} = 0.51 \text{ \AA}$ was determined from the individual normal temperature factors²⁰, $B_i = 2(2\pi\delta_{N_i})^2$. The approximation of uniform displacement amplitudes was made to simplify the simulation of the normal and variational scattering.

Figure 2c shows the simulated variational scattering calculated from a simple, liquid-like model for the short-range coupling of the atomic movements. The intensity distribution for variations of type S can be represented as

$$I_{V_S}(\mathbf{R}) = [1 - e^{-(2\pi\delta_S\mathbf{R})^2}] \text{FT} \sum P_{V_S}(\mathbf{r}). \quad (1)$$

The bracketed term is the complement of the Debye-Waller function for movements of r.m.s. amplitude δ_S , and it signifies that intensity which is lost from the normal Bragg scattering due to disorder should appear in the variational scattering⁹⁻¹². The second term is the Fourier transform (FT) of the sum of the autocorrelation functions $P_{V_S}(\mathbf{r})$ of the independently moving domains of type S , and $\mathbf{r}$ denotes the inter-atomic vectors. We surmise that the sum of the autocorrelation functions of the short-range coupled domains corresponds to $P_0(\mathbf{r})$, the Patterson function of the ideally-ordered unit-cell structure, (that is $\text{FT}[F_0(\mathbf{R})]^2$), for inter-atomic vectors less than some coupling distance γ_S and vanishes for longer vectors. Therefore we write $\sum P_{V_S}(\mathbf{r}) = P_0(\mathbf{r})T(\mathbf{r}/\gamma_S)$ where $T(\mathbf{r}/\gamma_S)$ is a monotonically decreasing function, such as a gaussian $e^{-(1/2)(\mathbf{r}/\gamma_S)^2}$, or an exponential $e^{-(\mathbf{r}/\gamma_S)}$, which smoothly truncates the Patterson function. Thus the variational scattering $I_{V_S}(\mathbf{R})$ can be approximated in terms of the ideally-ordered crystal structure and two adjustable parameters δ_S and γ_S .

Figure 3 illustrates how the intersection of the sphere of reflection with the symmetrical, three-dimensional scattering function

$$I_{V_S}(\mathbf{R}) = [1 - e^{-(2\pi\delta_S\mathbf{R})^2}] \text{FT}[P_0(\mathbf{r}) e^{-(1/2)(\mathbf{r}/\gamma_S)^2}] \quad (2)$$

generates the strikingly asymmetric variational scattering pattern shown in Fig. 2c. The parameters chosen to simulate the data in Fig. 1c are $\delta_S = 0.4 \text{ \AA}$ for the mean displacement amplitude and $\gamma_S = 6 \text{ \AA}$ for the mean coupling distance. Replacing the gaussian in equation 2 with an exponential decay having the same $6 \text{ \AA}$ mean coupling distance sharpened the pattern corresponding to Fig. 2c but did not alter the peak positions. From the spacings of the fringes in the variational scattering (Fig. 1c), the estimated spread of values for γ_S is $\approx 4-8 \text{ \AA}$. Within these short-range coupling distances, the linked internal movements could include torsion of backbone segments and neighbouring side-chain displacements.

Variational scattering, I_{V_L} , due to coupling of rigid body molecular movements in the crystal lattice over a range γ_L with r.m.s. amplitude δ_L can be represented by using these parameters instead of γ_S and δ_S in equation 2 and replacing $P_0(\mathbf{r})$ with $P_S(\mathbf{r}) * L(\mathbf{r})$ (that is, the Patterson function of the average unit-cell structure with internal fluctuations, δ_S , convoluted with the real-space lattice). As $\text{FT}[(P_S(\mathbf{r}) * L(\mathbf{r}))T(\mathbf{r}/\gamma_L)] = |F_0(hkl)|^2 e^{-(2\pi\delta_S R_{hkl})^2} * \text{FT} T(\mathbf{r}/\gamma_L)$, this term represents the convolution of the Bragg reflections $|F_S(hkl)|^2$ with the Fourier transform of the truncating function $T(\mathbf{r}/\gamma_L)$ (for example, $\text{FT}[e^{-(1/2)(\mathbf{r}/\gamma_L)^2}] = \sqrt{2\pi}\gamma_L e^{-(1/2)(2\pi\gamma_L\Delta R)^2}$, where ΔR is the distance from each reciprocal lattice point as origin). Thus, the shape of the haloes around the Bragg reflections should correspond to the Fourier transform of the function which defines the range γ_L over which the displacements δ_L are coupled. For $\gamma_L \gg$ unit cell dimensions, adjacent haloes will not overlap and the integrated intensity of each halo of index h ($= hkl$) will be given by $[1 - e^{-(2\pi\delta_L R_h)^2}] |F_0(h)|^2 e^{-(2\pi\delta_S R_h)^2}$ (which is equal to

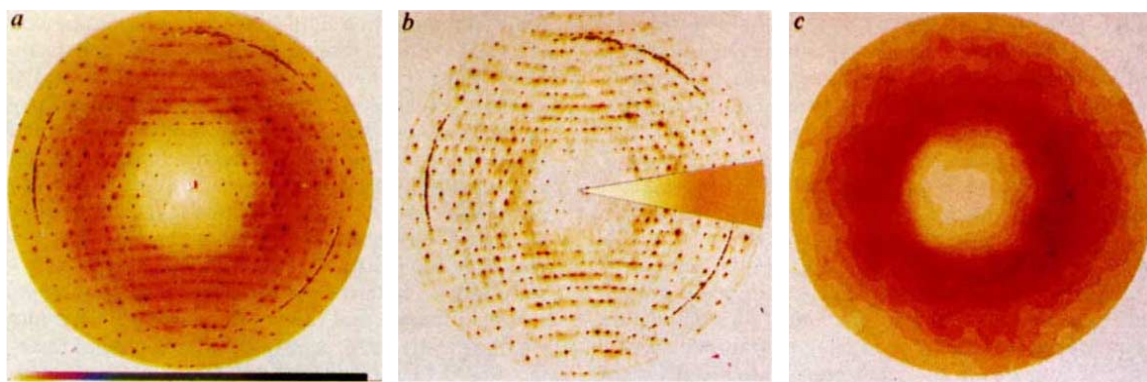


Fig. 1 X-ray diffraction data from insulin. *a*, A 40-h exposure of a stationary crystal (3-fold axis nearly vertical) recorded with focused, monochromatic Cu $K_{\alpha 1}$ radiation from an Elliott rotating anode. The data (corrected for camera background and polarization) are displayed out to radius $R = 0.45 \text{ \AA}^{-1}$. Bragg reflections are overexposed. The sharp arcs are due to diffraction from the Al foil window of the He beam tunnel. The colour table (optical density range 0–2 OD units) was constructed to distinguish small variations in intensity up to 0.5 OD units. *b*, Bragg reflections and haloes digitally separated by subtracting the smoothly-varying diffuse scattering component from the film data. The inset wedge shows the estimated circularly symmetrical Compton-plus water scattering. *c*, Variational scattering evaluated from the difference between *a* and the two components in *b* and *c* is $1.5 \times$ that in *a*. Each intensity step in *c* equals 0.02 OD units in the data.

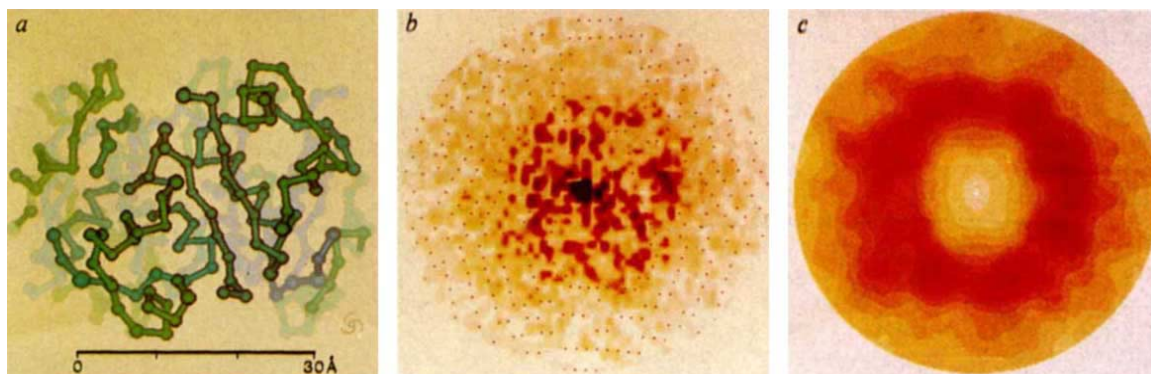
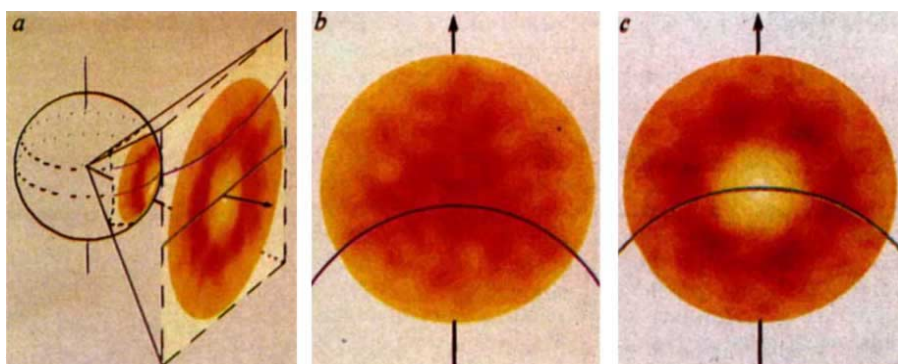


Fig. 2 Unit cell structure and model data. *a*, Backbones of the six insulin molecules in the unit cell, as viewed along the X-ray beam (A chain, green; B chain, blue). Depth cueing distinguishes the three dimers connected by β -strands. Sphere radii are $1.5 \times$ the standard deviations of the thermally agitated α -carbon atom electron density distributions. *b*, Calculated positions of the reciprocal lattice points are superimposed on the normal unit-cell scattering function, $|F_N(\mathbf{R})|^2$, plotted for the refined crystal orientation. Normal fringe widths are inversely proportional to the dimensions of the hexamer ($\approx 34 \text{ \AA} \times 47 \text{ \AA}$). *c*, Simulation of the variational scattering from a model of the molecules in the unit cell with internal movements of r.m.s. amplitude $\delta_S = 0.4 \text{ \AA}$ which are coupled over a gaussian range $\gamma_S = 6 \text{ \AA}$. Variational fringe widths, calculated from the model, are inversely proportional to γ_S . The intensity is displayed on the same step scale as in Fig. 1c.

Fig. 3 Geometry of data simulation. *a*, Intersection of calculated scattering function, $I_{vs}(\mathbf{R})$, with sphere of reflection (radius = $1/1.54 \text{ \AA}$), and projection of pattern (radius = $1/2.2 \text{ \AA}$) on film plane, to produce Fig. 2c. Vertical line through the sphere marks the crystal 3-fold axis. Arrow denotes the X-ray beam direction. Dashed circles mark equatorial and upper level sections. *b*, Section at $z = 1/4 \text{ \AA}$ perpendicular to the 3-fold axis of the scattering function, showing the intersection of the sphere (blue) which projects at the blue arc in the film plane. *c*, Equatorial section has 6-fold symmetry with quasi-mirror lines in the plane. Intersection of sphere (green) samples $I_{vs}(\mathbf{R})$ asymmetrically because the quasi-2-fold axis of the crystal is not aligned with the X-ray beam.



$|F_S(h)|^2 - |F_N(h)|^2$ if δ_S and δ_L are the only components defining the normal temperature factor).

According to the Waller theory^{10,11}, the intensity distribution of thermal diffuse scattering due to lattice vibrations in an isotropically elastic crystal should fall off as $1/\Delta R^2$ with distance ΔR from each reciprocal lattice point. In contrast to this prediction, we observe a much more gradual fall-off in intensity of the haloes in the insulin diffraction pattern (Fig. 1b). Fitting

the haloes by a gaussian or a lorentzian, we estimate a coupling range $\gamma_L \approx 20\text{--}30 \text{ \AA}$, which corresponds to the size of an insulin monomer or dimer. This limited range suggests that lattice vibrations are over-damped and that the movements of the molecules in the crystal are only significantly correlated between nearest neighbours, as in a liquid. Liquid-like local coupling of movements in two-dimensional crystals of polystyrene latex spheres in water has been directly observed by microscopy²¹.

These colloidal crystals may provide a better analogy for the coupling of movements in protein crystals than perfectly elastic lattices (such as diamonds), in which long-range coupling of atomic movements can be represented in terms of normal modes, as described by the theories of Debye^{9,22} and Waller¹⁰.

Our measurements indicate that there are two principal independent components, δ_S and δ_L , to the atomic movements in the insulin crystal. If these are the only components then $\delta_S^2 + \delta_L^2 = \delta_N^2 = (0.51 \text{ \AA})^2$ and, as $\delta_S \approx 0.4 \text{ \AA}$ (compare with Fig. 2c), δ_L should be $\sim 0.3 \text{ \AA}$. The value of δ_L , which represents rigid-body displacement of the molecules, cannot be greater than the smallest net atomic displacements, δ_N , which are $\sim 0.25 \text{ \AA}$, as measured from the individual temperature factors²⁰. An independent estimate of δ_S and δ_L can be made from the integrated intensities of the variational scattering because, from equation 2 and the corresponding relation for I_{V_L} , $\int I_{V_S}/\int I_{V_L} \approx \delta_S^2/\delta_L^2$. The integrated intensity I_{V_S} due to short-range variations (Fig. 1c) is three to four times larger than $\int I_{V_L}$ measured from the haloes (Fig. 1b), after correcting for the sharp Bragg reflections. With $(\delta_S/\delta_L)^2 = 3.5$, and $\delta_N = 0.51 \text{ \AA}$, the two unique components would be $\delta_S = 0.45 \text{ \AA}$ and $\delta_L = 0.24 \text{ \AA}$. Values of δ_S from 0.4–0.45 $\text{\AA}$ produce similar simulations of the very diffuse protein scattering (Fig. 2c), which account for 60–80% of the total variational scattering. Thus, most of the variational scattering appears to be due to movements that are coupled over distances in the ranges $\gamma_S \approx 4\text{--}8 \text{ \AA}$ and $\gamma_L \approx 20\text{--}30 \text{ \AA}$.

From a simple mechanical model for thermal fluctuations in a protein crystal, Morazov and Morazova²³ have shown that the mean-square, rigid-body displacements of the molecules in the lattice can be approximated by $\delta_L^2 = kT/2Ea$, where E is Young's modulus, a is the molecular diameter and $kT/2$ is the thermal energy. From our X-ray scattering measurements on insulin crystals, $\gamma_L \approx a$ and $\delta_S^2/\delta_L^2 \approx \gamma_L/\gamma_S$, which implies that $\delta_S^2 \approx kT/2E\gamma_S$. Thus, we infer that the microscopic ratio of stress to strain for the short-range deformations of the backbone and side-chain packing within each molecule may be comparable to the macroscopic Young's modulus for distortions of the crystal lattice.

It is remarkable that a simple model for liquid-like coupling of displacements in a protein crystal, defined by two adjustable parameters that are compatible with other measurements of movement, can provide an approximate representation of the variational scattering due to fluctuations both within the molecule and between molecules in the lattice. All atoms in the protein do not, however, have the same amplitude displacement. Local differences in the r.m.s. amplitudes δ_N for movements of the α -carbon atoms measured by the individual temperature factors²⁰ are illustrated by the different size spheres in Fig. 2a. The variational scattering we have calculated from a model for the internal movements, using individual mean-square atomic displacements $\delta_S^2 = \delta_N^2 - \delta_L^2$, with $\delta_L = 0.25 \text{ \AA}$, is similar to that shown in Fig. 2c for the uniform displacement model. We infer that this similarity results because the different amplitude movements are not segregated in domains with distinctly different structure. Detailed comparisons with model scattering functions, to determine the individual components of the displacements and the coupling relations, will require mapping the variational scattering in three dimensions. Preliminary measurements of the intensity distribution, using a high spatial resolution, TV-based X-ray area detector²⁴, are compatible with the model calculations (for example, Fig. 3).

Molecular dynamics simulations²⁵ of fluctuations in proteins may be tested by comparing the calculated variational scattering (that is, $\langle |F_i(\mathbf{R})|^2 \rangle - \langle |F_i(\mathbf{R})|^2 \rangle$ from the Fourier transforms of the time-varying models) with measurements of the X-ray scattering due to internal movements. The prediction, based on the temporal correlations of simulated fluctuations, that the interior of

a protein should be fluid-like²⁶ is confirmed by our analysis of the short-range spatial correlations of movements inside insulin. Models for the molecular movements calculated from normal modes²⁷, which assume undamped elastic coupling, should show correlations on the scale of the largest molecular dimensions; therefore, it is unlikely that such models would fit our X-ray scattering data. Although there is no evident long-range coupling of the internal movements, these movements can be influenced by lattice interactions. Comparisons that we have made, in collaboration with W. C. Phillips and R. M. Sweet, of the variational scattering from triclinic and tetragonal lysozyme crystals indicate that the amplitudes of internal movements are smaller in the more tightly packed triclinic lattice than in the tetragonal form, which correlates with the differences in temperature factors^{28,29} and viscoelastic properties²³ of these crystals. We anticipate that the liquid-like internal movements of protein molecules may be represented by models of local regions whose mechanics are constrained (but not uniquely determined) by their environment.

We thank W. C. Phillips and R. M. Sweet for advice, J. Li and G. N. Phillips Jr for stimulating insights, and R. Oldenbourg and X. Wen for helpful information. This work was supported by the National Cancer Institute.

Received 13 November 1987; accepted 17 February 1988.

- Phillips, G. N., Fillers, J. P. & Cohen, C. *Biophys. J.* **32**, 485–502 (1980).
- Boylan D. & Phillips, G. N. *Biophys. J.* **49**, 76–78 (1986).
- Doucet, J. & Benoit, J. P. *Nature* **325**, 643–646 (1987).
- Artymiuk, P. *Nature* **325**, 575–576 (1987).
- Li, J. thesis, Harvard Univ. (1978).
- Maulik, S., Caspar, D. L. D., Phillips, W. C. & Goodenough, D. A. *Biophys. J.* **49**, 342a (1986).
- Maulik, S. thesis, Brandeis Univ. (1986).
- Caspar, D. L. D. & Holmes, K. C. *J. Molec. Biol.* **46**, 99–133 (1969).
- Debye, P. *Ann. d. Physik* **43**, 49–95 (1914); *The Collected Papers of Peter J. W. Debye* (Interscience, New York 1954).
- Waller, I. *Z. Phys.* **17**, 398–408 (1923).
- James, R. W. *The Optical Principles of the Diffraction of X-rays* (Cornell University Press, 1965).
- Amoros, J. L. & Amoros, M. *Molecular Crystals: Their Transforms and Diffuse Scattering*. (Wiley New York, 1968).
- Adams, M. J. *et al. Nature* **224**, 491–495 (1969).
- Compton, A. H. & Allison, S. K. *X-rays in Theory and Experiment* second edn (Van Nostrand, New York 1935).
- Salunke, D. M., Veerapandian, B., Kodandapani, R. & Vijayan, M. *Acta crystallogr.* **B41**, 431–436 (1985).
- Narten, A. H. *Oak Ridge National Laboratory Report No. 4578* (1970).
- Sakabe, N., Sakabe, K. & Sasaki, K. *Proc. int. Symp. Biomolec. Struct. Interactions, Suppl. J. Biosci.* Vol. 8 numbers 1 and 2, 45–55 (1985).
- Wonnacott, A. J., Brook, P. *Imperial College Rotation Film Scanning Package* (1984).
- Bernstein, T. F. *et al. J. Molec. Biol.* **112**, 535–42 (1977).
- Dodson, E. J., Dodson, G. G., Hodgkin, D. C. & Reynolds, C. D. *Can. J. Biochem.* **57**, 469–479 (1979).
- Murray, C. A. & Van Winkle, D. H. *Phys. Rev. Lett.* **58**, 1200–1203 (1987).
- Debye P., *Ann. d. Physik* **39**, 789–839 (1912).
- Morozov, V. N. & Morozova, T. Y. *A. J. theor. Biol.* **121**, 73–88 (1986).
- Kalata, K. *Meth. Enzym.* **114**, 486–510 (1985).
- Karplus, M. & McCammon, J. A. *CRC crit. Rev. Biochem.* **9**, 293–349 (1981).
- McCammon, J. A., Gelin, B. R. & Karplus, M. *Nature* **267**, 585–590 (1977).
- Levitt, M., Sander, C. & Stern, P. J. *molec. Biol.* **181**, 423–447 (1985).
- Ramanadham, M., Sieker, L. C., Jensen, L. H. & Birknes, B. J. *Acta Crystallogr.* **A37** C-33 (1981).
- Sternberg, M. J., Grace, O. E. P. & Phillips, D. C. *J. molec. Biol.* **130**, 231–253 (1979).

Erratum

Is the G18.95–1.1 radio source a shell-type supernova remnant with a central object?

A. R. Patnaik, T. Velusamy & V. R. Venugopal

Nature **332**, 136–137 (1988)

In this letter, line 11 should read “in G357.7–0.1 is identified with a compact HII region, and there”.

DNA footprinting with hydroxyl radical

T.D. Tullius

High resolution images of proteins bound to DNA can be achieved through the use of a simple inorganic chemical system that produces the hydroxyl radical.

DNA is a rather passive partner in the intricate choreography of genetic systems. It is only through the action of proteins that the messages encoded in DNA are read and acted upon. Of central importance to molecular biology, then, is the study of DNA-protein complexes. At the most basic level, identifying what protein binds to which DNA sequence will enable a better understanding of the mechanisms controlling gene expression, DNA replication, and all other biological processes which involve DNA. At a deeper level, achieving the ultimate goal of understanding "genetic switches" in chemical detail will require structural information on these DNA-protein complexes.

Answers to these questions have come from an experimental approach dubbed DNA "footprinting" which creates images of areas of protein binding to DNA. These images are produced by subjecting a DNA-protein complex to the action of a reagent which degrades DNA. The protein protects the part of the DNA to which it is bound from the reagent, while the rest of the DNA is cleaved. The image of the complex is revealed through electrophoresis of the DNA on a high-resolution polyacrylamide gel. Sites protected by the protein show up as blank spaces on the autoradiograph of the gel, and look like the "footprints" of the protein on the DNA. The footprinting technique can be applied to femtomole quantities of DNA-protein complexes, so that the experiments of the molecular biologist are not limited by the amount of material available.

The classic reagent used in DNA footprinting is the enzyme deoxyribonuclease I (DNase I)¹. This reagent is not ideal, however, because DNase I does not cleave uniformly along the DNA molecule. Countless images produced using DNase I have appeared in the literature over the past decade showing DNA-protein binding sites. The use of DNase I has allowed the basic identification of DNA-binding proteins and the sequences to which they bind, but structural information has remained elusive.

Radical chemistry

To better understand the structural interactions between DNA and DNA-binding proteins, my laboratory has pursued a new approach to footprinting that relies on the chemistry of a very small but extremely reactive molecule, the hydroxyl radical ($\bullet\text{OH}$)². This chemistry produces

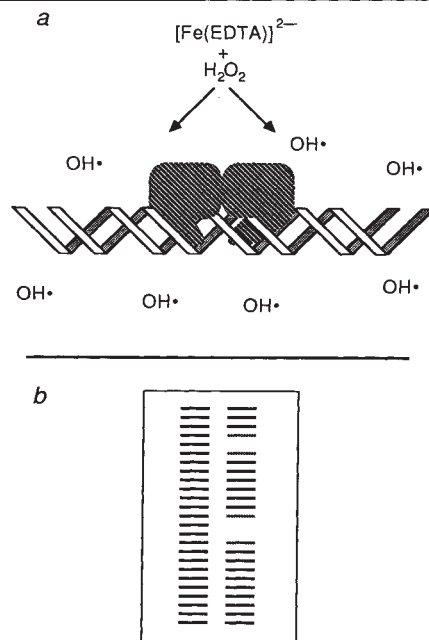
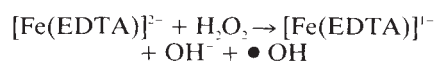


Fig. 1 Schematic diagram of the hydroxyl radical footprinting experiment. *a*, $[\text{Fe}(\text{EDTA})]^{2-}$, repelled from DNA by its negative charge, reacts at a distance from the DNA molecule with hydrogen peroxide to produce the hydroxyl radical. Bound protein protects the DNA backbone from cleavage by the radical. *b*, Footprints are analysed by gel electrophoresis. The left-hand lane shows bands of equal intensity produced by the cleavage of naked DNA by the hydroxyl radical. In the right-hand lane, several bands are lost in the cleavage pattern of the DNA-protein complex.

the highest resolution footprints available, while using inexpensive, readily available chemicals.

The hydroxyl radical has long been known to have the ability to break the backbone of DNA. Much of the mutagenic damage to DNA caused by ionizing radiation derives from the hydroxyl radicals produced by the interaction of high energy photons with water in the cell. Work in the laboratories of Haseltine³ and Dervan⁴ pointed out more recently that the reaction of iron(II) with O_2 is an excellent source of hydroxyl radicals for DNA cleavage. My laboratory has modified these schemes⁵ by using the EDTA complex of iron(II) in a reaction with hydrogen peroxide to generate the hydroxyl radical:



Sodium ascorbate may be added to the reaction to reduce the iron(III) EDTA product back to iron(II), creating a cata-

lytic cycle which requires only a micromolar concentration of iron.

Because the EDTA complex of iron possesses a negative charge, the metal complex does not bind to the polyanionic DNA molecule². Only the neutral hydroxyl radical reaches the DNA helix, yielding an extremely delicate probe for the surface of the DNA molecule.

A bound protein blocks the cleavage of DNA by hydroxyl radical, so that the image of the protein bound to DNA shows up as blanks in the electrophoresis pattern^{6,7} (Fig. 1). Two qualities of the hydroxyl radical give these footprints much higher structural resolution than is available using conventional footprinting reagents. First, because the promiscuous hydroxyl radical lacks base or sequence preference²⁻⁵, the specific nucleotides of the DNA in contact with the protein may be elucidated⁶. Second, the very small size of the hydroxyl radical allows it to react with open sites in the DNA-protein complex from which other cleavage reagents are excluded^{6,7}. The hydroxyl radical is, in effect, a reactive water molecule which provides a map of the accessibility to solvent of the DNA when it is bound to protein, yielding the chemical analogue of the calculation and display of solvent-accessible surfaces by molecular graphics computer systems.

This strategy was first applied to a pair of well-known DNA-protein complexes — the interaction of the bacteriophage lambda repressor and cro proteins with the O_R1 operator sequence⁸. The hydroxyl radical footprints of the repressor- and cro-DNA complexes show small patches of protection three or four nucleotides long, separated by five to seven efficiently cleaved nucleotides. The patterns on the two strands are identical in shape, but offset from each other in the 3' direction by three nucleotides (Fig. 2). The hydroxyl radical footprints demonstrate that these proteins bind to one side of the DNA helix, an observation which substantiates the extensive crystallographic⁹ and other¹⁰ data on these systems.

The binding of the archetypal zinc finger protein TFIIIA to DNA has also been studied using this footprinting technique, by carrying out the reaction directly in a wheat germ extract for a series of deletion mutants of TFIIIA produced by *in vitro* translation¹⁰. These deletion mutants were designed to eliminate one finger after another from each end of the protein molecule. Computer analysis of

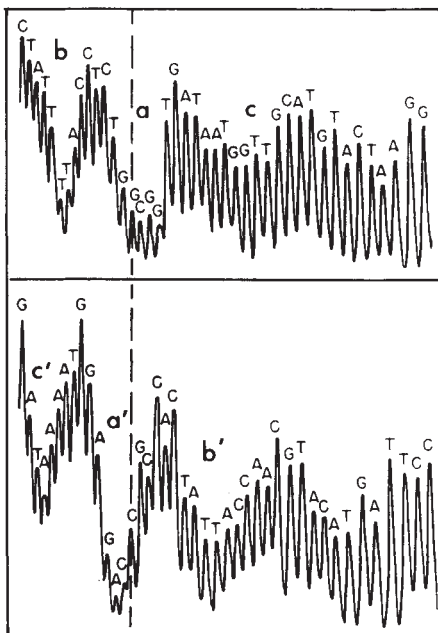


Fig. 2 Densitometric scans of hydroxyl radical footprints of both strands of DNA (aligned vertically by base pairs) of the bacteriophage lambda repressor bound to the O_R1 operator sequence. The broken vertical line passes through the pseudo-dyad axis of the operator sequence, showing the three-nucleotide offset of the footprints of the two strands. The unprotected nucleotides between footprints *a* and *b* show that the protein binds to one side of the DNA molecule.

densitometric scans of footprints of the various deletion mutants result in maps of the contacts with DNA of six of the nine fingers of TFIIIA. The high resolution and reproducibility of hydroxyl radical footprints reward detailed analysis by densitometry of the sometimes subtle footprint patterns on the autoradiograph. Much of the information contained in hydroxyl radical footprints is discarded if the gels are analysed visually.

Refinements in the method will allow more information to be extracted from hydroxyl radical footprints. My laboratory plans to use the X ray structure of the lambda repressor-DNA cocrystal¹¹ to calibrate the structural information in hydroxyl radical footprints. □

Thomas D. Tullius is at the Department of Chemistry, The Johns Hopkins University, Baltimore, Maryland 21218, USA. For more information, fill in reader service number 100.

- Galas, D.J. & Schmitz, A. *Nucleic Acids Res.* **5**, 3157-3170 (1978).
- Tullius, T.D. *Trends biochem. Sci.* **12**, 297-300 (1987).
- D'Andrea, A.D. & Haseltine, W.A. *Proc. natn. Acad. Sci. U.S.A.* **75**, 3608-3612 (1978).
- Hertzberg, R.P. & Dervan, P.B. *Biochemistry* **23**, 3934-3945 (1984).
- Tullius, T.D. & Dombroski, B.A. *Science* **230**, 679-681 (1985).
- Tullius, T.D. & Dombroski, B.A. *Proc. natn. Acad. Sci. U.S.A.* **83**, 5469-5472 (1986).
- Tullius, T.D., Dombroski, B.A., Churchill, M.E.A. & Kam, L. *Meth. Enzym.* **155**, 537-558 (1987).
- Pabo, C.O. & Lewis, M. *Nature* **298**, 443-447 (1982).
- Johnson, A.D., Ph.D. dissertation, Harvard Univ. (1980).
- Vrana, K., Churchill, M.E.A., Tullius, T.D. & Brown, D.D. *Molec. cell. Biol.* (in the press).
- Jordan, S.R., Whitcombe, T.V., Berg, J.M. & Pabo, C.O. *Science* **230**, 1383-1385 (1985).

Analytica highlights

Some of Europe's finest products for the scientific market will be shown at the Analytica exhibition next week in Munich. Several are highlighted below.

ON stand 2C25, Russell pH Ltd will be displaying selections from its range of **ion selective electrodes**, pH meters, and kits (*Reader Service No. 101*). Russell's ion selective electrodes have detachable cable electrodes, and measure directly the concentration of both cations and anions. The electrodes also measure dissolved gases such as ammonia, sulphur dioxide and carbon dioxide. The model CD660 pH electrode is a dedicated bench pH/ion meter geared toward the needs of analytical laboratories. The model CD660 has a 12 mm LED display which gives readings accurate to 0.001 pH, 0.1 mV and 0.1 °C, says Russell. Full slope and buffering facilities and recorder output are included. Kits from Russell include the KnipHe kit for the pH measurement of meat and semi-solid foods, and a new low-conductivity water test kit for the analysis of water samples.

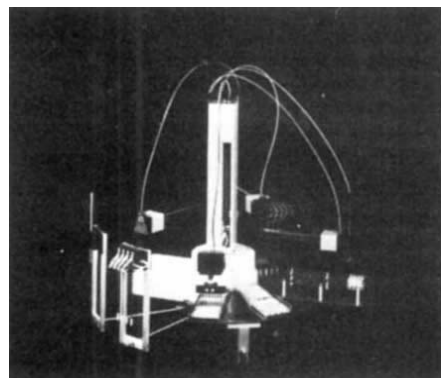
Trivector's **colour measurement systems** can be seen for the first time in Europe on stand 2B20 (*Reader Service No. 102*). Trivector says the model CL6000 systems employ a new design of optics and microprocessor, and are priced for low budgets. A four-element filter set mounted in a mechanically and thermally stable housing enables them to yield results with



Trivector's colour measurement system comes in configurations for the research lab to the shop floor.

a reproducibility and stability of greater than 1 per cent, says the company. Reflectance and transmittance heads for solids or liquids are available; the transmittance head is designed for cuvettes of up to 50 mm. Software is available for transmitting data from the colour measurement systems to a PC or mainframe computer. Trivector makes four levels of colour measurement systems, configured for research to industrial applications.

The Kemble Instrument Company will be introducing a **laboratory robotics system**, the Autolab 500, on stand 1B11 (*Reader Service No. 103*). The Autolab 500 is operated using standard IBM-compatible software which contains built-



Kemble Instrument Company's Autolab 500 laboratory robot has a long reach.

in error handling facilities. It can lift a load of 3 kg and move it at a speed of greater than 2.5 ms⁻¹, says Kemble. The Autolab 500 uses a central rotating column and a robotic arm to access a circular working area. The radial approach offers a large and expandable working area and allows for interaction with other laboratory equipment. The robot can run third-party software packages for the control of tube and plate readers, incubator/shakers and balances. Kemble offers a range of tools for the robotic system, including sample probes, tube washers, plate washers, tip pick-up mechanisms and grippers for both tubes and microtitre plates.

Paesel's new **biochemicals catalogue**, available on stand 3A3, covers products for use in research, diagnosis or scale-up (*Reader Service No. 104*). In its thirteenth edition, the Paesel catalogue lists 2,800 products by their alphabetical names and by a simplified categorical index. More than 100 eicosanoids, prostaglandins and leukotrienes are included, plus over 1,000 antisera, purified and labelled antibodies, and monoclonal antibodies to a large number of antigens. Biological agents such as peptides, agonists and inhibitors for neuroscience, receptor studies and pharmacological research are also described in the catalogue, as are pure enzymes, fluorescent markers or dyes for a variety of labelling techniques. If the substance desired is not in the catalogue, Paesel offers a service to provide biochemicals by special order.

Heraeus, exhibiting on stand 1B9, has an alternative to gassing with ethylene oxide or treatment with ionizing radiation for the **sterilization** of heat-labile or consumer goods as they enter and leave

sterile areas (*Reader Service No. 105*). The Ozone Lock model SD5050 is a hatch system which uses UV radiation and ozone gas for disinfection. Heraeus says the Ozone Lock system is more economical and safer than other gassing methods, because it contains a catalyst which decomposes the ozone gas within 15 minutes. Most goods and equipment can be sterilized in 2-12 hours with the Ozone Lock.

The small world

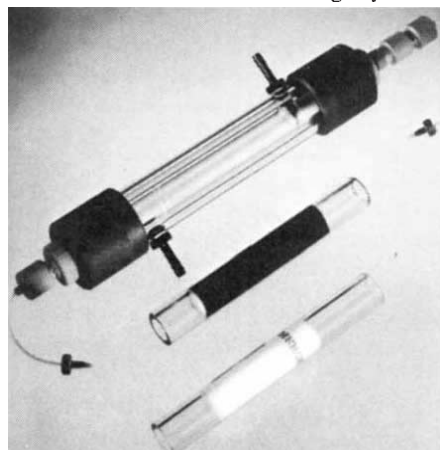
Bio-Rad Microscience is launching an optical system for converting standard research microscopes into **laser scanning confocal microscopes**, on stand 3B14 (*Reader Service No. 106*). The new MRC-500 system is optimized for the needs of the biomedical scientist, and functions in either reflected light or fluorescence mode. The mirror scanning system produces images with a one frame per second refresh rate and provides the capability to zoom the magnification beyond the normal range of an optical microscope, says Bio-Rad. Weak signals may be enhanced by a number of filtering modes, including Kalman integration. The system is provided with a multi-line argon ion laser, and the transfer system is fully achromatic at all wavelengths from UV to near infra-red. The MRC-500 is controlled by a desktop computer with cards for controlling scanning and image acquisition. The software allows two images to be displayed simultaneously, or to be merged for direct comparison.

Analytical Measuring Systems will be displaying its full line of **image analysis systems** on the AI Tektron stand, number 16D17 (*Reader Service No. 107*). Analytical Measuring Systems will be featuring the Optomax V, its fully automatic image analyser, which has the capability to measure a wide range of parameters including area, diameter, feature count and longest dimension. The Optomax V processes all data during the course of analysis using IBM and PS/2 compatible software which performs statistics, histograms and cumulative frequency distributions. When combined with the company's Quickstep automatic scanning

and focusing system, Analytical Measuring Systems says the Optomax V provides unattended analysis.

Chromatography currents

Merck has a new line of inert, biocompatible and pressure-resistant **glass chromatography columns** (*Reader Service No. 108*). The Superformance range of glass columns are based on a cartridge system:

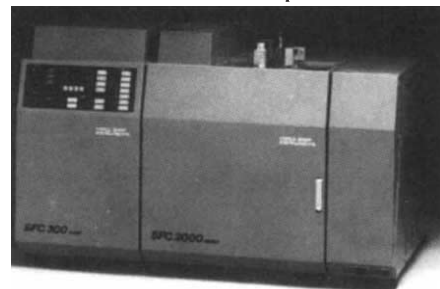


See-through chromatography columns mean interchangeability and no dead volume to Merck columns can easily be removed from the column holder and replaced with a different column for another purpose. Merck says the columns offer no dead volume during sample introduction, because the substance is homogeneously distributed to the whole face of the column directly from the infeed capillary. The columns are suitable for HPLC; the 10 mm i.d. columns can tolerate a pressure of up to 120 bar. Merck will be exhibiting on stand 3A16.

Hewlett-Packard recently introduced what may be the smallest mass spectrometer on the market with the launch of its model HP 5971A mass selective detector (MSD) (*Reader Service No. 109*). Designed for GC/MS analyses, the new MSD mounts on the side of the HP 5890A gas chromatograph, providing classical electron impact spectra while taking up just seven linear inches of bench space. Hewlett-Packard says total ion scanning and selective ion monitoring may be car-

ried out with the new MSD, and the system is also optimized for capillary chromatography. The HP 5971A will scan any mass range from 10-650 AMU, at any of eight selectable scan rates. Both scan rates and mass ranges may be programmed over time, allowing close examination of masses of interest. Each eluting peak may be scanned repeatedly over the chosen mass range; resulting spectra are stored and may be displayed as a total ion chromatogram or as individual spectra for each scan. Hewlett-Packard is exhibiting on stand 7A9.

On stand 2A18, Carlo Erba will be exhibiting the SFC 3000 — a **supercritical fluid chromatograph** which also offers the capabilities of a high temperature capillary gas chromatograph (*Reader Service No. 110*). The basis of the SFC 3000 is an HT Mega oven that can be operated at temperatures of up to 500 °C, and that is equipped with two separate injectors, columns and detectors. The portion of the SFC 3000 dedicated to supercritical fluid



Carlo Erba's super-critical fluid chromatograph.

chromatography has an automatic time-controlled injector, supercritical fluid chromatography column and outlet restrictor, and uses the full range of gas chromatography detectors. The supercritical fluid is delivered by a syringe pump which can be programmed to operate at constant pressure, constant flow, or other specified mode.

Proteins and peptides

Boehringer Mannheim Biochemicals, exhibiting on stand 7B7, has a new range of **sequencing-grade proteases** (*Reader Service No. 111*). Boehringer Mannheim says the proteases are submitted to a stringent set of quality control procedures. To ensure purity, all are checked by either reversed phase HPLC or SDS gel electrophoresis/silver staining. All proteases are also checked for specificity using a model protein substrate of known amino acid sequence. An 18-hour digestion using a 10-fold excess of enzyme is performed to show no non-specific cleavage as analysed by reverse phase HPLC. Boehringer Mannheim claims the proteases are very active at high dilutions, and are active in denaturing agents such as SDS, guanidine HCL and urea. The proteases are packed as lyophilized aliquots,

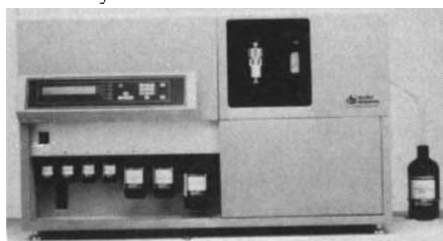


Analytical Measuring Systems will be exhibiting its full line of image analysis systems, including the Optomax V.

and come with a complete description of buffers and reaction conditions.

Biometra has shortened the time required for the transfer of large molecules to blotting membranes with its Biometra-Fastblot **electroblotting instrument** (*Reader Service No. 112*). Biometra says the Biometra-Fastblot performs most blotting procedures in several minutes, and that electroblotting of nucleic acids for hybridization purposes can be accomplished within one to two hours. The Biometra-Fastblot has a cooling option to prevent the denaturation of proteins and nucleic acids under the high electrical current. A safety system protects the user from electrical accident, even if the instrument is connected to the power supply. For currents of 50-2,000 mAmp, Biometra offers a small, stabilized power pack. Biometra will be exhibiting on stand 9A12.

Applied Biosystems will be announcing the availability of a new **simplified protein/peptide sequencer** at its stand, number 3B10 (*Reader Service No. 113*). The model 471A sequencer uses Edman chemistry and includes a built-in PTH



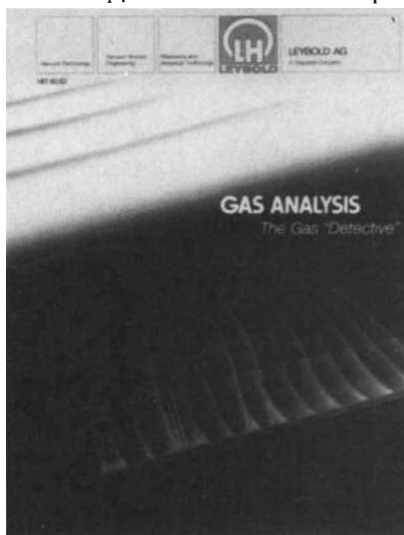
Applied Biosystems' new protein/peptide sequencer focuses on convenience and economy.

amino acid separation system for convenience and economy. The protein sequencer has a patented valve block design which is incorporated in over 2,000 Applied Biosystems instruments worldwide. The model 471A uses TFA to cleave

the PTC peptide, and recycles the PTH separation solvent. Applied Biosystems says the model 471A protein sequencer can sequence as little as 250 pmol of protein with 92 per cent or greater repetitive yield.

Spectroscopy news

Leybold AG, exhibiting on stand 11C9, has a new **brochure on gas analysis** that describes applications for its line of quad-



The full line of quadrupole mass spectrometers for gas analysis are described in Leybold AG's brochure.

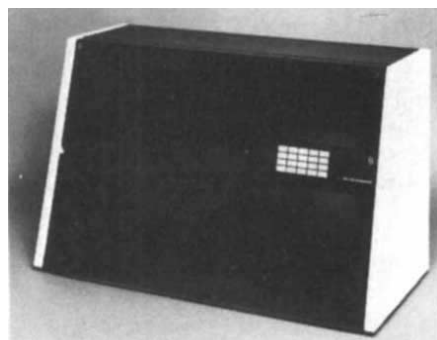
rupole mass spectrometers (*Reader Service No. 114*). The model QAS 100 mass spectrometer from Leybold is a portable system with an integrated computer equipped with an emergency power unit. Leybold's model MSQ 1002 mass spectrometer comes on a 19-inch trolley, and is especially suited for process optimization during development, the surveillance of different, non-continuously monitored processes, and the routine maintenance of vacuum-technology systems. The model MSQ 1003 is a modular mass spectrometer for stationary applications such as monitoring coating processes, metallurgical processes, or gas analysis for the use of waste gas. Each of Leybold's instruments has a mass range of 1-99 AMU. The QAS 100 detects gases down to a concentration of 1 p.p.m., and the MSQ 1002 and MSQ 1003 are sensitive down to 10 p.p.m.

Hamamatsu offers various high-speed **photomultiplier tubes** with rise times from subnanoseconds to a few nanoseconds (*Reader Service No. 115*). These tubes feature high gain and are well suited for fluorescence lifetime measurements and other pulse-counting applications, according to the company. The tubes employ specialized devices to minimize false signals such as ringing and afterpulse. The tubes come in one side-on and two head-on configurations, with tube diameters of between one and two inches. Spectral response for the side-on tube ranges from

185 to 900 nm, and for the head-on tubes from 300 to 650 nm. Hamamatsu will be exhibiting in stand 12A16.

Chelsea Instruments will be introducing a **Fourier transform spectrometer** designed to work in the UV and visible region on its stand, number 16B16 (*Reader Service No. 117*). Chelsea Instruments' model FT500 was developed and designed at Imperial College in London and modified by the company. Chelsea Instruments claims the instrument offers a resolution of 0.03 cm^{-1} , and throughput that is one order of magnitude better than that of a grating spectrometer of similar resolution. The Fourier transform is performed on a dedicated Analogic AP500 array processor capable of transforming 1×10^6 complex points in a matter of minutes. The data is digitally acquired and spectral processing is performed using highly sophisticated custom-designed software.

Applied Photophysics has announced the release of new **data acquisition and analysis software** for use with its kinetic spectrometers (*Reader Service No. 116*). Its stopped-flow instrument makes time-resolved absorption and emission measurements following rapid reagent mixing. The data system available for the stopped-flow instrument runs on an IBM



The new Fourier transform spectrometer from Chelsea Instruments functions in the UV and visible region.

PC/AT/XT with a 12-bit ADC interface and signal conditioning amp. Program selections can be made with either menus or function keys. Three methods for determining infinity reading are provided, including the Swinbourne method, which is used when infinity reading is drifting due to a subsequent slow reaction. The software calculates first-order rate constants by linear or non-linear least squares analysis, and displays reaction curves graphically. Applied Photophysics will be exhibiting on stand 1B7. □

These notes are compiled by Carol Ezzell from information provided by the manufacturers. To obtain further details about these products, use the reader service card bound inside the journal. Prices quoted are sometimes nominal and apply only within the country indicated.

SINGER
INSTRUMENTS

**Cell
Injection?**

**Singer MK III
Singer Microforge**

Singer Instrument Co. Ltd.

Treborough Lodge
Roadwater, Watchet
Somerset, England, TA23 0QL
Tel: Washford (0984) 40226
Telex: 337492 Comcab G

Reader Service No.50